Journal of Medical Internet Research

Journal Impact Factor (JIF) (2023): 5.8
Volume 21 (2019), Issue 11 ISSN 1438-8871 Editor in Chief: Gunther Eysenbach, MD, MPH

Contents

Viewpoints

Reviews

Original Papers

Commentary

Discussion Papers

Corrigenda and Addendas

Viewpoint

From a Digital Bottle: A Message to Ourselves in 2039

Alejandro R Jadad[1], MD, DPhil; Tamen M Jadad Garcia[2], BSc

[1]Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada
[2]Beati Inc, Toronto, ON, Canada

Corresponding Author:
Alejandro R Jadad, MD, DPhil
Dalla Lana School of Public Health
University of Toronto
155 College St, 6th Floor
Suite 2404
Toronto, ON, M5T 3M7
Canada
Phone: 1 4163585631
Email: a.jadad@utoronto.ca

Abstract

We are fully aware that we could have wasted our time writing this message, as nobody might read it. Even those who read it might ignore it, and those who read and care about it might be unable to do anything. It may simply be too late. Nevertheless, this message describes the hopes we had back in 1999, imagining how the incredible digital tools whose birth we were witnessing, could change the world for the better. In 2019, when we wrote these words, we were saddened to realize that most of what we had imagined and proposed in the past 20 years could have been written the day before, without losing an iota of relevance. Whoever or whatever you might be, dear reader—a human, a sentient machine, or a hybrid—we would like you to understand that, rather than an attempt to predict the future, which probably continues to be an impossible endeavor, this message was meant to act as an invitation, regardless of when or where it is found, to engage in a conversation that has already transcended time and space, even if the issues it contains have become irrelevant.

(J Med Internet Res 2019;21(11):e16274) doi:10.2196/16274

KEYWORDS

wisdom; pandemic; concepts; future; extinction; self-sabotage; precariat; stupidity; noosphere; capitalism

To whom might read this, human or sentient machine:

We are fully aware that we could have wasted our time writing this message, as nobody might read it. Even those who read it might ignore it, and those who read and care about it might be unable to do anything. It may simply be too late.

We would like to share with you the hopes we had back in 1999, imagining how the incredible digital tools whose birth we were witnessing, could change the world for the better. As a Babyboomer father, who, as a physician, had started communicating with patients via email in 1990 and a Millennial daughter who saw ways to improve healthcare since her infancy [1], we have spent the last two decades dreaming about and proposing new possibilities that would enable everyone to enjoy a healthy life.

It saddens us to realize that most of what we had imagined and proposed in the past two decades could have been written yesterday, without losing an iota of relevance today.

Twenty years ago, we imagined new forms of partnerships between healthcare professionals and the public [2-4], enlightened by trustworthy information, available anywhere, in the right format, in the right amount, with the right balance [5-8]. We called for a major shift "from our ethic of competition and narrow self interest, focused on gadgets—to one of generosity and collaboration, centred on people." [9]

Instead, we are now facing an avalanche of "fake" or misleading knowledge that is capable of reaching the masses more effectively than verified facts, particularly through social media [10]. This has opened the door to new spaces for hate speech and manipulation of public opinion, undermining trust and hindering efforts to control diseases that should have been eradicated already, or that could otherwise be managed easily [10-12].

In the past two decades, we dreamed about ways in which we could use every new technological development to create tailored services, available to all people, that could transcend

our traditional geographic, cultural, religious, institutional, and political boundaries [13-15]. We believed that it was possible to join forces across the world to imagine and reinvent how we care for each other, through the enthusiastic and proficient use of information and communication technologies [16,17].

Instead, in 2019, we face a landscape littered by variations in the same disconnected examples that have been used for two decades to illustrate how digital tools could contribute to the prevention, diagnosis, or treatment of diseases. Sadly, the promises of information and communication technologies to transform healthcare services remain unfulfilled as a result of poorly designed interfaces, persistently increasing workload for providers, unresolved privacy concerns, lack of sensible models for the reimbursement of new services, chronically outdated communication skills, and obsolete rules of etiquette to guide interactions at all levels [18,19].

We also imagined new possibilities to unlearn, to "un-see" and "un-believe" in the barriers that prevent us from having a full life, from before our first breath to the last one [20]. We dreamed that we could look at death as the main source of insights that could inspire us to live fully, without fear, regrets or distress, through all stages of our lives [21-23].

Instead of broadening our understanding of the mental and social aspects of our lives, we kept steering our technological innovations to go deeper and deeper into the body. Since the turn of the 20th century, informatics has enabled an explosion in the biological understanding of what ails us, bringing back the dreams of yore about our capacity to personalize medical responses to physical challenges. Such insights, nevertheless, have also emboldened the medical industrial complex to boost its efforts to create sophisticated weapons to feed the war against diseases. The infatuation around "fixing" physical problems is now manifesting through an apparently inexhaustible menu of ever-reductionist "omics" [24,25], which are taking the chemical-mechanical view of humanity to new depths [26] while feeding a frantic new era of hypermedicalization of life. Instead of being a source of humility and insights, death has now become something for which there might be a "cure" [27].

During the past two decades, we used online platforms to nurture a global conversation about the meaning of health [28,29]. This led to a new conceptualization that views it as the ability, which anyone could develop, to adapt and manage the inevitable challenges we face through life as individuals or communities [30]. We invited the world to unleash a pandemic of health through the enlightened and generous use of digital technologies [31,32]. We viewed the internet as a treasure trove of opportunities to celebrate and boost the "high-touch" with the "high-tech" [16].

Rather than using them to usher in truly planet-wide, open, equitable, affordable services that are culturally sensitive and tailored to our unique needs and those of other living beings, sophisticated digital platforms are accelerating the privatization and segmentation of a "sickcare" system in practically every region of the world while transforming it into a branch of the financial system [33]. The same phenomenon of "financialization," driven by global digital platforms and algorithms, has engulfed all other aspects of human life, compounding the threats to health at all levels, from the individual to the planetary level. Digital technologies are at the core of the engines that have made 2019 a year full of nefarious records in terms of global warming, major weather events, deforestation, and extinction of animal and plant species [34-37]. As a result, environmental factors have become the main cause of one quarter of all deaths in the world [38].

We hoped, above all, that digital technologies would enable us to prevail, transforming us into "humanodes" in a global superorganism, feeding a Noosphere—a planetary thinking network of reasoning minds—in which we all could divert our seemingly unavoidable course toward a world without humans to one in which we would thrive [39].

We were naive…

The opportunities we have had over the last 20 years appear to have been "insurmountable."

Instead of realizing the dream of a humanity that could be connected deeply and meaningfully into an enlightened whole, technology over the past 20 years has been used, relentlessly, to dismantle communities and societies around the world, segmenting them all the way to the individual level [40]. This extreme level of "precision capitalism," which is fueled by social media, is leading to unprecedented levels of isolation [41], making loneliness a growing source of preventable deaths [42] and low levels of well-being [43].

Our growing disconnection from each other is also facilitating the emergence of new economic models ruled by a progressively smaller group of unaccountable masters who have unleashed a growing network of all-powerful algorithms. Such models consider the vast majority of humans either as consumers of the goods or services they peddle or as biological automatons that are relevant only because they are capable of performing tasks that machines are still unable to complete or because they are less expensive [44].

The relegation of humans to the margins of the economy is breeding fresh threats to health, which multiply and reinforce each other as technology takes over every aspect of human life. A case in point is the impact that automation and artificial intelligence are having on existing occupations and on the ways in which humans make a living. As machines and artificial intelligence gain dominance, a new class of unfortunate humans is emerging, under the guise of freedom [45]. Known collectively as "the precariat," these are people who are self-employed or freelancers, doing mostly "gig jobs" without long-term or permanent contracts [46]. Often, they are overqualified individuals who work under short-term or zero-hour contracts, sometimes for a few minutes at a time, without benefits provided by employers [47]. Even physicians are now at risk of joining this class [48].

With the pervasive use of digital technology driving people into precariousness, marketeers, engineers, and scientists at the service of the new unchecked overlords have also devised new tricks to extract the remaining disposable income available to the helpless masses. In 2019, people's attention has become the main target [49]. As a result, digital devices—or, more precisely, the experiences they purvey—have been manipulated to become

increasingly addictive, thanks to constant monitoring and experimentation on consumers through the platforms they use and the data they produce [50]. This has generated new problems, such as those associated with gaming, gambling, or over-consumption of social media, and in the invigoration of old ones, manifested through shopaholism, workaholism, eating disorders, online pornography, or the rapid proliferation of electronic cigarettes [51].

In 2019, the new threats created by the massive digitalization of the planet are finding all those charged with safeguarding or improving human health wanting. Front line healthcare providers, in particular, are ill equipped to face the vastly complex array of new challenges. In the past 20 years, instead of becoming examples of how to lead healthy lives, physicians and nurses are facing high levels of depression, stress, anxiety, and burnout [52-55]. The prevalence of addiction is likely similar or higher among physicians when compared with the general public [56]. They also commit suicide more often, with women at an even greater risk [57]. In fact, suicide is the only cause of death that has a higher prevalence among physicians than the general population. Male physicians are 40% more likely than members of the general public to kill themselves, while the risk is more than double for female physicians [58]. The situation among nurses is likely similar [59,60].

As we kill ourselves in ever more creative ways, the ability of machines to control those who will remain alive is increasing. As the so-called "The Internet of Things" materializes, a practically unimaginable number of interconnected devices, compounded by our cavalier ways to share our information with corporations, governments and financial institutions is curtailing our freedom. Insidiously, we are becoming prisoners in a digital panopticon run by a surveillance capitalist apparatus [61]. Given that it is estimated that the number of connected computing devices will exceed one trillion in the next 20 years [62], we are facing a world in which instead of humans controlling the Internet of Things, we could become one of the many things controlled by the internet [63].

We can no longer afford to be oblivious to what is obvious.

The space for humanity to flourish in the middle of a spectrum spanning from apocalyptic and transhumanist extremes has narrowed significantly in the past 20 years [64]. We might be facing the worst possible scenario: a self-inflicted apocalyptic posthuman era, brought about by our seemingly unstoppable propensity to self-sabotage.

Instead of using our digital offspring to nurture the Noosphere, we have created the "Atesphere"—after Ate, the goddess of delusion, blind folly, and ruin [65]—an interconnected planetary layer of human stupidity, fed by our self-harming tendencies. Within it, we are hastening our complete or near-complete extinction.

To counterbalance the risk that such a pessimistic stance could lead to a self-fulfilling prophecy, we would like to put forward a counter argument [66]: It is also possible for humans and machines to create a "Sophosphere," a planetary network of interconnected people, machines, and all other living things striving to negate self-sabotage through wisdom.

We believe that wisdom could be the best antidote to self-sabotage because, based on a conceptualization proposed in 1999 [67], we view it as "the ability to know what it is to live well, and do everything possible to achieve this, given the circumstances, while enabling others to do the same, as a means to overcome the destructive mental processes and behaviours that inhibit our own ability to experience a full life, in our own way, together." This approach, which we feel adds balance to the Atesphere, makes it possible for machines to be wise too and for machines and humans to collaborate and create the conditions needed to live well in harmony with other living things.

In a consistently mirror-like fashion, the Sophosphere might occur, as a Big Bang, in response to the devastating effects of the Atesphere at its peak. Just as it happens with individuals who undergo a near-death experience, it has been suggested that our species could go through a similar process, snapping back from the brink of death, en masse, to embrace radically different priorities and values and be ready to live fully [68].

Another option, also counterintuitive, involves doing nothing, on the basis of the belief that our frantic efforts to make change happen might be part of the problem, rather than part of the solution. Instead, this approach calls for restraint and patience, as it regards massive social change as a nonlinear phenomenon that could happen suddenly, in response to the accumulation of little adjustments made by large enough numbers of individuals or small groups to their own lives [69].

In addition, the Sophosphere could emerge more deliberately, through the collective efforts of innovators from different sectors, joining forces in and from different regions of the world to incubate new ways to coexist harmoniously, which could then spread, through social contagion, until they become pandemic [32].

As noted a few years ago, the most sensible option might be hardest to accept, as it requires recognizing that humanity is terminally ill as the result of self-inflicted wounds, and that the wisest course of action is to engage in palliative care measures at a species level [39].

Having reached this point, dear reader, whoever or whatever you might be, we would like you to understand that, rather than an attempt to predict the future, which is impossible, this message that was released into the cyberwaves in 2019 should be viewed as an invitation, regardless of when or where it is found, to engage in a conversation that has already transcended time and space, even if the issues it contains have become irrelevant.

We will only know what will happen when it happens.

Conflicts of Interest

None declared.

Multimedia Appendix 1
[PNG File , 1602 KB - jmir_v21i11e16274_app1.png]

References

1. Jadad AR. A view from the Internet age: let's build a health system that meets the needs of the next generation. CMAJ 2004 Dec 07;171(12):1457-1458 [FREE Full text] [doi: 10.1503/cmaj.1041276] [Medline: 15583185]
2. Jadad AR. Promoting partnerships: challenges for the internet age. BMJ 1999 Sep 18;319(7212):761-764. [doi: 10.1136/bmj.319.7212.761]
3. Jadad AR. I am a good patient, believe it or not. BMJ 2003 Jun 12;326(7402):1293-1295. [doi: 10.1136/bmj.326.7402.1293]
4. O'Grady L, Jadad A. Shifting from Shared to Collaborative Decision Making: A Change in Thinking and Doing. J Particip Med 2012;34(19):1608-1616.
5. Jadad AR, Gagliardi A. Rating Health Information on the Internet. JAMA 1998 Feb 25;279(8):611. [doi: 10.1001/jama.279.8.611]
6. Gagliardi A, Jadad A. Examination of instruments used to rate quality of health information on the internet: chronicle of a voyage with an unclear destination. BMJ 2002 Mar 09;324(7337):569-573 [FREE Full text] [doi: 10.1136/bmj.324.7337.569] [Medline: 11884320]
7. Jadad A. The Internet and evidence-based decision-making: a needed synergy for efficient knowledge management in health care. J Med Internet Res 2000 Sep 13;2:e2. [doi: 10.2196/jmir.2]
8. Eysenbach G, Jadad AR. Evidence-based patient choice and consumer health informatics in the Internet age. J Med Internet Res 2001 Jun 7;3(2):E19 [FREE Full text] [doi: 10.2196/jmir.3.2.e19] [Medline: 11720961]
9. Jadad AR, Delamothe T. What next for electronic communication and health care? BMJ 2004 May 13;328(7449):1143-1144. [doi: 10.1136/bmj.328.7449.1143]
10. Finn P. The Impact of Social Media on Communication Sciences and Disorders: A Need for Examination and Research. Perspect ASHA SIGs 2019 Apr 15;4(2):224-227 [FREE Full text] [doi: 10.1044/2019_pers-st-2019-0001]
11. Shoup JA, Narwaney KJ, Wagner NM, Kraus CR, Gleason KS, Albright K, et al. Social Media Vaccine Websites: A Comparative Analysis of Public and Moderated Websites. Health Educ Behav 2019 Jun 29;46(3):454-462. [doi: 10.1177/1090198118818253] [Medline: 30596265]
12. Armstrong PW, Naylor CD. Counteracting Health Misinformation. JAMA 2019 May 21;321(19):1863. [doi: 10.1001/jama.2019.5168]
13. Jadad A, Goel V, Rizo C, Hohenadel J, Cortinois A. The global e-health innovation network - Building a vehicle for the transformation of the health system in the information age. Business Briefing: Next Generation Healthcare 2000;48:54.
14. Cortinois A, Downey S, Closson T, Jadad A. Hospitals in a globalized world: a view from Canada. Healthc Pap 2003;4(2):14-32. [Medline: 14660881]
15. Shachak A. Electronic Health Records in the Age of Social Networks and Global Telecommunications. JAMA 2010 Feb 03;303(5):452. [doi: 10.1001/jama.2010.63]
16. Balliol College, University of Oxford. 2007. An innovator in global healthcare and human well-being URL: https://www.balliol.ox.ac.uk/alumni-and-friends/floreat-domus/2007/an-innovator-in-global-healthcare-and-human-well-being [accessed 2019-10-15]
17. Jadad AR, Serra M, Palacio D, Espinal S, Rodriguez D, editors. Trusted networks: The Key to Achieve World-class Health Outcomes on a Shoestring. Toronto, ON: Beati Inc; 2018.
18. The Lancet. 2019 Apr 04. Barriers and Facilitators for Implementation of Electronic Consultations (eConsult) to Enhance Access to Specialist Care: A Scoping Review URL: https://papers.ssrn.com/abstract=3365087 [accessed 2019-10-15]
19. Schraeder T. Physician Communication: Connecting with Patients, Peers, and the Public. Oxford, UK: Oxford University Press; 2019.
20. Jadad AR. Unlearning: Incomplete Musings On The Game Of Life And The Illusions That Keep Us Playing. Toronto, ON: Foresight Press; 2008.
21. Enkin M, Jadad AR, Smith R. Death can be our friend. BMJ 2011 Dec 21;343(dec21 2):d8008-d8008. [doi: 10.1136/bmj.d8008]
22. Herrera-Molina E, Librada S, Lucas M, Jadad-Garcia T, Rodriguez S, Jadad A. The New Health Foundation: Transforming palliative care through the integration of clinical and social services, and community engagement. Eur J Palliat Care 2017;24:123-125.
23. Herrera E, Jadad-Garcia T, Librada S, Alvarez A, Rodriguez Z, Lucas M, et al, editors. Beginning from the End: Enabling the transformation of how we care for each other by bringing together the power of health and social services, and the community, during the last stages of life. Seville, Spain: New Health Foundation; 2017.

24. Lu Y, Chang Y, Hoffman E, Yu G, Herrington D, Clarke R, et al. Integrated Identification of Disease Specific Pathways Using Multi-omics data. bioRxiv 2019 Jun 11 [FREE Full text] [doi: 10.1101/666065]
25. Deshmukh A. Gene and disease: an ?omics? as the pillars of systems. World J Pharm Res 2019;8(6):216.
26. Glouberman S. The Mechanical Patient: Finding a More Human Model of Health. Milton Park, UK: Productivity Press; 2018.
27. Spivak J. HeinOnline. 2019. Redefining Life: The Constitutional Implications of Patenting Immortality URL: https://heinonline.org/HOL/LandingPage?handle=hein.journals/caelj37&div=35&id=&page= [accessed 2019-10-28]
28. Jadad AR, O'Grady L. The BMJ opinion. 2008 Dec 08. A global conversation on defining health URL: https://blogs.bmj.com/bmj/2008/12/10/alex-jadad-on-defining-health/ [accessed 2019-10-15]
29. Jadad AR, O'Grady L. How should health be defined? BMJ 2008 Dec 10;337(dec10 1):a2900-a2900. [doi: 10.1136/bmj.a2900]
30. Huber M, Knottnerus JA, Green L, van der Horst H, Jadad AR, Kromhout D, et al. How should we define health? BMJ 2011 Jul 26;343(jul26 2):d4163-d4163. [doi: 10.1136/bmj.d4163] [Medline: 21791490]
31. Jadad AR. Creating a pandemic of health: What is the role of digital technologies? J Public Health Policy 2016 Nov 29;37(Suppl 2):260-268. [doi: 10.1057/s41271-016-0016-1] [Medline: 27899800]
32. Jadad AR, Arango A, Sepúlveda J, Espinal S, Rodriguez D, Wind K, editors. Unleashing a pandemic of health from the workplace: Believing is seeing. Toronto, ON: Beati Inc; 2017.
33. Hunter BM, Murray SF. Deconstructing the Financialization of Healthcare. Development and Change 2019 Jun 06;50(5):1263-1287. [doi: 10.1111/dech.12517]
34. Graham F. Nature. 2019 Aug 21. Daily briefing: July was the hottest month ever recorded on Earth URL: http://www.nature.com/articles/d41586-019-02536-1 [accessed 2019-10-15]
35. Gibbens S. National Geographic. 2019. The Amazon is burning at record rates—and deforestation is to blame URL: https://www.nationalgeographic.com/environment/2019/08/wildfires-in-amazon-caused-by-deforestation/ [accessed 2019-10-15]
36. Lewis M. Electrek. 2019 Sep 14. Climate Crisis Weekly's extreme weather breaks records URL: https://electrek.co/2019/09/14/climate-crisis-weekly-2019-extreme-weather-breaks-records-un-climate-change-threatens-human-rights/ [accessed 2019-10-15]
37. IPBES. 2019. Nature's dangerous decline 'unprecedented,' species extinction rates 'accelerating' URL: https://www.ipbes.net/news/Media-Release-Global-Assessment [accessed 2019-10-15]
38. The Global Environment Outlook. 2019. UNEnvironment URL: https://www.unenvironment.org/interactive/global-environment-outlook/ [accessed 2019-10-15]
39. Jadad AR, Enkin M. Does humanity need palliative care? European Journal of Palliative Care 2017 May:102-103.
40. Rushkoff D. Team Human. In: Team Human. New York, NY: WW Norton & Company; 2019:256.
41. Twenge JM, Spitzberg BH, Campbell WK. Less in-person social interaction with peers among U.S. adolescents in the 21st century and links to loneliness. Journal of Social and Personal Relationships 2019 Mar 19;36(6):1892-1913. [doi: 10.1177/0265407519836170]
42. Williams SE, Braun B. Loneliness and Social Isolation—A Private Problem, A Public Issue. Journal of Family & Consumer Sciences 2019 Feb 01;111(1):7-14. [doi: 10.14307/jfcs111.1.7]
43. Twenge JM. More Time on Technology, Less Happiness? Associations Between Digital-Media Use and Psychological Well-Being. Curr Dir Psychol Sci 2019 May 22;28(4):372-379. [doi: 10.1177/0963721419838244]
44. Berardi F. Futurability: The Age of Impotence and the Horizon of Possibility. Brooklyn, NY: Verso; 2019.
45. Conen W, Schippers J. Self-Employment as Precarious Work A European Perspective. In: Self-employment: between freedom and insecurity. Cheltenham, UK: Edward Elgar Publishing; 2019.
46. Raskin J. CounterPunch. Mar 19. Guy Standing on Anxiety, Anger and Alienation: an Interview About The Precariat URL: https://www.counterpunch.org/2019/03/19/guy-standing-on-anxiety-anger-and-alienation-an-interview-about-the-precariat/ [accessed 2019-10-15]
47. Godlewska-Bujok B, Patulski A. Precarious Work. In: Precariat: next stage of development or economic predominance in a new scene?. Cheltenham, UK: Edward Elgar Publishing; 2019.
48. Shah NR. Health Care in 2030: Will Artificial Intelligence Replace Physicians? Ann Intern Med 2019 Feb 26;170(6):407 [FREE Full text] [doi: 10.7326/m19-0344]
49. Franck G. The economy of attention. Journal of Sociology 2018 Nov 27;55(1):8-19. [doi: 10.1177/1440783318811778]
50. Berthon P, Pitt L, Campbell C. Addictive De-Vices: A Public Policy Analysis of Sources and Solutions to Digital Addiction. Journal of Public Policy & Marketing 2019 Jul 28;38(4):451-468. [doi: 10.1177/0743915619859852]
51. Choi J, King DL, Jung Y. Editorial: Neurobiological Perspectives in Behavioral Addiction. Front Psychiatry 2019 Jan 22;10:3-3 [FREE Full text] [doi: 10.3389/fpsyt.2019.00003] [Medline: 30723426]
52. Williams E, Rathert C, Buttigieg S. The Personal and Professional Consequences of Physician Burnout: A Systematic Review of the Literature. Med Care Res Rev 2019 Jun 19:1077558719856787. [doi: 10.1177/1077558719856787] [Medline: 31216940]
53. Hoff T, Carabetta S, Collinson GE. Satisfaction, Burnout, and Turnover Among Nurse Practitioners and Physician Assistants: A Review of the Empirical Literature. Med Care Res Rev 2019 Feb 13;76(1):3-31. [doi: 10.1177/1077558717730157] [Medline: 28901205]

54. Yu F, Raphael D, Mackay L, Smith M, King A. Personal and work-related factors associated with nurse resilience: A systematic review. Int J Nurs Stud 2019 May;93:129-140. [doi: 10.1016/j.ijnurstu.2019.02.014] [Medline: 30925279]
55. Petrie K, Crawford J, Baker STE, Dean K, Robinson J, Veness BG, et al. Interventions to reduce symptoms of common mental disorders and suicidal ideation in physicians: a systematic review and meta-analysis. The Lancet Psychiatry 2019 Mar;6(3):225-234. [doi: 10.1016/s2215-0366(18)30509-1]
56. Angie CC, Leung T. Substance Use Disorders. In: Weiss Roberts L, Trockel M, editors. The Art and Science of Physician Wellbeing: A Handbook for Physicians and Trainees. Berlin, Germany: Springer International Publishing; 2019.
57. Dutheil F, Aubert C, Pereira B, Dambrun M, Moustafa F, Mermillod M. Lancet. 2019 Jun 07. Suicide Among Physicians and Health-Care Workers: A Systematic Review and Meta-Analysis URL: https://papers.ssrn.com/abstract=3397193 [accessed 2019-10-15]
58. Albuquerque J, Tulk S. Physician suicide. CMAJ 2019 May 06;191(18):E505-E505. [doi: 10.1503/cmaj.181687] [Medline: 31061076]
59. Davidson J, Proudfoot J, Lee K, Zisook S. Arch Psychiatr Nurs. 2019 Jun 8. Nurse suicide in the United States: Analysis of the Center for Disease Control 2014 National Violent Death Reporting System dataset URL: http://www.sciencedirect.com/science/article/pii/S0883941719300287 [accessed 2019-10-15]
60. Mumba M, Kraemer K. Substance Use Disorders among Nurses in Medical-Surgical, Long-Term Care, and Outpatient Services. Medsurg Nurs 2019;28(2):118.
61. Zuboff S. The age of surveillance capitalism: The fight for a human future at the new frontier of power. New York, NY: PublicAffairs; 2019.
62. Anonymous. The Economist. 2019. How the world will change as computers spread into everyday objects URL: https://www.economist.com/leaders/2019/09/12/how-the-world-will-change-as-computers-spread-into-everyday-objects [accessed 2019-10-15]
63. Hendricks VF, Vestergaard M, editors. Epilogue: Digital Roads to Totalitarianism. In: Reality Lost: Markets of Attention, Misinformation and Manipulation. Berlin, Germany: Springer International Publishing; 2019:119.
64. Jadad AR, Enkin MW. Computers: transcending our limits? BMJ 2007 Jan 06;334(suppl_1):s8-s8. [doi: 10.1136/bmj.39038.663970.94]
65. ATE. URL: https://www.theoi.com/Daimon/Ate.html [accessed 2019-10-15]
66. Benoit L, Russo T, Barry C, Falissard B, Henckes N. "You have to believe in something": Risk of psychosis and psychiatrists' beliefs in the self-fulfilling prophecy. Soc Sci Med 2019 Jun;230:20-29. [doi: 10.1016/j.socscimed.2019.03.035] [Medline: 30947102]
67. Ryan S. What is wisdom? Philos Stud 1999;93(2):39.
68. Dossey L. Possibilities for the Survival of the Human Race: A Global Near-Death Experience. Explore (NY) 2018 Jul;14(4):241-247. [doi: 10.1016/j.explore.2018.04.001] [Medline: 29859823]
69. Poland B. Coming Back to Our True Nature: What Is The Inner Work that Supports Transition? In: Zywert K, Quilley S, editors. Health in the Anthropocene. Toronto, ON: University of Toronto Press; 2019.

Edited by G Eysenbach; submitted 15.09.19; peer-reviewed by M Focsa; comments to author 11.10.19; revised version received 16.10.19; accepted 17.10.19; published 01.11.19.

Please cite as:
Jadad AR, Jadad Garcia TM
From a Digital Bottle: A Message to Ourselves in 2039
J Med Internet Res 2019;21(11):e16274
URL: https://www.jmir.org/2019/11/e16274
doi:10.2196/16274
PMID:31682578

Review

Electronic Health Self-Management Interventions for Patients With Chronic Kidney Disease: Systematic Review of Quantitative and Qualitative Evidence

Hongxia Shen[1], MSc; Rianne M J J van der Kleij[1,2], PhD; Paul J M van der Boog[3], MD, PhD; Xinwei Chang[4], MSc; Niels H Chavannes[1], MD, PhD

[1]Department of Public Health and Primary Care, Leiden University Medical Centre, Leiden, Netherlands

[2]Department of Obstetrics and Gynaecology, Erasmus Medical Center, Rotterdam, Netherlands

[3]Department of Nephrology, Leiden University Medical Centre, Leiden, Netherlands

[4]Department of Surgery, School of Nutrition and Translational Research in Metabolism, Maastricht University, Maastricht, Netherlands

Corresponding Author:
Hongxia Shen, MSc
Department of Public Health and Primary Care
Leiden University Medical Centre
Albinusdreef 2
Leiden, 2333 ZA
Netherlands
Phone: 31 633789207
Email: H.Shen@lumc.nl

Abstract

Background: Chronic kidney disease (CKD) poses a major challenge to public health. In CKD patients, adequate disease self-management has been shown to improve both proximal and distal outcomes. Currently, electronic health (eHealth) interventions are increasingly used to optimize patients' self-management skills.

Objective: This study aimed to systematically review the existing evidence regarding the implementation and effectiveness of eHealth self-management interventions for patients with CKD.

Methods: Following a search in 8 databases (up to November 2017), quantitative and qualitative data on process and effect outcomes were extracted from relevant studies. Quality was appraised using the Crowe Critical Appraisal Tool; narrative synthesis was performed to analyze the data extracted.

Results: Of the 3307 articles retrieved, 24 (comprising 23 studies) were included in this review; of these, almost half were appraised to be of low to moderate quality. There was considerable heterogeneity in the types of interventions used and the outcomes measured. A total of 10 effect and 9 process outcome indicators were identified. The most frequently reported effect outcome indicators were specific laboratory tests and blood pressure (BP), whereas satisfaction was the most frequently reported process outcome indicator. Positive effects were found for proximal outcomes (eg, BP control and medication adherence), and mixed effects were found for more distal outcomes (eg, quality of life). High feasibility, usability, and acceptability of and satisfaction with eHealth self-management interventions were reported. The determinant ability of health care professionals to monitor and, if necessary, anticipate on patient measurements online was mostly cited to influence patients' adherence to interventions.

Conclusions: eHealth self-management interventions have the potential to improve disease management and health outcomes. To broaden the evidence base and facilitate intervention upscaling, more detailed descriptions and thorough analysis of the intervention components used are required. In addition, our review reveals that outcomes closely related to the scope and duration of the intervention implemented are most likely to be impacted. For instance, if a 4-week Web-based training to optimize disease management skills is implemented, the outcome perceived control would more likely be affected than kidney function. Although this seems obvious, most studies evaluate only distal outcomes and thereby fail to capture intervention effects that might contribute to long-term health improvement. We advise future researchers to carefully consider their choice of outcomes based on their sensitivity for change. In this way, we ensure that relevant effects are captured and legitimate conclusions are drawn.

(J Med Internet Res 2019;21(11):e12384) doi:10.2196/12384

KEYWORDS

eHealth; self-management; systematic review; chronic kidney disease

Introduction

Background

Chronic kidney disease (CKD) is a major public health concern [1-3]. Globally, more than 70 million individuals are affected by CKD [4]. CKD is defined as kidney damage or a measured glomerular filtration rate (GFR) of ≤60 mL/min/1.73m^2 for more than 3 months. CKD is classified into 5 stages based on GFR decline [5]. The level of kidney function deterioration has a direct relationship with an increase in morbidity and mortality [6], poorer patient outcomes [3], higher hospitalization rates [7], and substantial increase in health care expenditures [8]. Patients with CKD report a lower quality of life (QoL) [9] and may experience severe medical complications and cognitive dysfunction [10].

Disease self-management (hereafter referred to as *self-management*) is defined as "an individual's ability to manage the symptoms, treatment, physical and psychosocial consequences, and lifestyle changes inherent to the life with a chronic condition" [11]. Adequate self-management is reported to improve patients' health behaviors targeted by the intervention (ie, proximal outcomes) and also indirect outcomes, such as disease characteristics and progress (ie, distal outcomes) [12-14]. Although the potential benefits of self-management interventions are widely reported in the literature, extrapolating these results in day-to-day practice is difficult. Lack of efficacy in practice might be related to a suboptimal implementation of the self-management interventions [15,16]. Reported barriers were often related to intervention characteristics, such as lack of tailoring to the individual patient. Moreover, a lack of patient involvement in intervention design and insufficient care continuity and accessibility were reported to hamper implementations [17,18].

Electronic health (eHealth) technologies can help address implementation barriers by making interventions more accessible, acceptable, tailored, and interactive [19-21]. The most cited definition of eHealth is that by Eysenbach [22]:

> *e-health is [...] referring to health services and information delivered or enhanced through the Internet and related technologies. In a broader sense, characterizes [...] to improve health care locally, regionally, and worldwide by using information and communication technology.*

eHealth can help patients achieve personal health goals, and it allows them to feel more responsible for their health status [23]. Moreover, eHealth facilitates remote patient communication and exchange of health data, helping to increase health care efficiency while maintaining a wide-scale, cost-effective health care approach [24]. eHealth interventions have been successfully implemented to support weight loss [25,26], promote smoking cessation [27], reduce depressive symptoms [28], and decrease mortality rates and acute admissions [29]. In addition, eHealth-based interventions have been successfully applied to manage chronic disease [30-32].

Several studies have reported the use of eHealth-based self-management interventions in CKD [33-36]. Moreover, 3 systematic reviews were published on this topic [37-39]. However, these reviews only concentrated on 1 particular eHealth application, such as telemedicine; dietary mobile apps; and automated information technology tools. Moreover, these reviews focused on a limited number of study designs and outcomes. For example, 2 reviews only included randomized controlled trials (RCTs) [38,39], and 1 review excluded studies focusing on implementation outcomes such as feasibility, validity, and acceptability [39]. Moreover, none of these reviews [37-39] reviewed the contribution of individual intervention components (eg, self-monitoring) to the effects found. These limitations of previous reviews make it difficult for researchers and intervention developers to determine which components should be employed to maximize the effectivity of eHealth self-management interventions for CKD patients.

Objectives

This study, therefore, aimed to systematically review the available evidence on eHealth-based self-management interventions for CKD. In specific, we aimed to review the following: (1) study characteristics and type of eHealth applications used; (2) intervention components implemented and, if possible, their relative contribution to the effect found; (3) both process and effect outcomes; and (4) determinants of implementation.

Methods

Protocol and Registration

This review was performed according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement [40]. The protocol was registered in the international Prospective Register of Systematic Reviews database (Centre for Reviews and Dissemination [CRD] number: CRD 420 180 81681).

Search Methodology

A systematic search was conducted to identify relevant articles; the search strategy was developed in collaboration with a certified librarian. In total, 8 electronic databases (PubMed, EMBASE, Web of Science, Cochrane Library, EmCare, PsycINFO, Academic Search Premier, and Science Direct) were searched in November 2017. Search terms covered 3 areas: (1) CKD, (2) eHealth, and (3) self-management (see Multimedia Appendix 1). Reference lists of the included studies were searched to identify other relevant articles. EndNote X9 (Clarivate Analytics) was used to support the review process.

Eligibility Criteria

Inclusion and exclusion criteria (Textbox 1) were determined using the Patients, Interventions, Comparison, Outcomes, Study design methodology [41].

Textbox 1. Inclusion and exclusion criteria for this study.

Inclusion criteria:

- Participants—patients classified with chronic kidney disease (stage 1-5)
- Intervention—eHealth technologies ("any information and communication technology designed to deliver or enhance health services and information") applied to facilitate chronic kidney disease patients' self-management ("the care taken by individuals towards their own health and well-being: it comprises the actions they take to lead a healthy lifestyle; to meet their social, emotional and psychological needs; to care for their long-term condition, and to prevent further illness or accidents") [11]
- Comparison—no restrictions
- Outcomes—articles reporting on clinical (ie, patients' intermediate outcomes or clinical parameters of disease severity, such as blood pressure, fluid management, and mortality), humanistic (ie, consequences of disease or treatment on patients' functional status or quality of life, such as physical functioning, well-being, and levels of depression or anxiety), economic and utilization (ie, measures of health resource utilization, medical costs, and cost-effectiveness), and/or process (ie, indicators that affect patient care by improving health care delivery or patient-health care interactions and self-management related–factors, such as adherence to intervention, usability of eHealth technologies, and self-efficacy) outcomes
- Language restrictions—articles needed to be written in English
- Study design—randomized and nonrandomized controlled trials, noncomparative trials, and qualitative or mixed methods articles

Exclusion criteria:

- Type of electronic health used—studies with devices only used for communication (eg, a telephone only used for a follow-up call) or data collection (eg, an internet system solely used to collect patient data without further intervention) purposes
- Study design—case reports containing ≤3 participants, commentaries, reviews, letters, dissertations, editorials, conference proceeding, and books

Study Identification

After removal of duplications, titles and abstracts of the retrieved articles were screened independently by 2 reviewers (HS and XC). Articles that did not meet inclusion criteria were removed. Potentially relevant articles were obtained in full text and reviewed independently by 2 authors (HS and XC). Any disagreements between the 2 authors were resolved by consensus or consultation with a third author (RK).

Data Collection

Data collection was performed independently by 2 reviewers (HS and XC) using a standardized data extraction form. Study characteristics, descriptions of eHealth self-management interventions (eg, intervention components), process and effect outcome indicators, and determinants of implementation were extracted. Discrepancies in extraction were discussed until consensus was reached.

Quality Assessment

Article quality was appraised independently by HS and XC using the Crowe Critical Appraisal Tool (CCAT) [42]—a reliable, widely used quality appraisal tool [43,44]. Use of the CCAT user guide promoted validity and inter-rater reliability [43-46]. The CCAT form is divided into 8 categories and 22 items, with a total of 99 subitems. Subitems are rated on a scale of *present*, *absent*, or *not applicable*. A 6-point scale ranging from 0 (the lowest) to 5 (the highest) is used to assign score per category, with 40 being the maximum achievable total score.

The CCAT does not allow for a qualitative comparison of appraisal scores. Hence, we used the star score system developed by our research group to compare study quality [47]. First, we calculated a quality score based on the CCAT. Then, a mean score and standard deviation of the quality scores were calculated. Star scores were then assigned to each article: 1 star if a quality score was more than 1 SD below mean; 2 stars if a quality score ranged from 1 SD below mean to mean score. The kappa between the 2 reviewers' scores of quality assessment was 0.63, reflecting substantial agreement [48].

Data Synthesis

Data were reviewed using narrative synthesis [49]. Study characteristics were reviewed, summarized, and analyzed in a spreadsheet. In accordance with previous categorizations of eHealth [32,39,50], eHealth self-management interventions were split into 5 major types (see Multimedia Appendix 2). eHealth functionalities used were described based on the technology functionality framework [51,52]. In addition, based on the operationalization by Mohr et al [53], eHealth-based self-management interventions included were further detailed: (1) intervention components (based on Morrison et al [54]; see Multimedia Appendix 3)—active intervention parts that support self-management behavior, including elements defined as *what* is provided to the user (eg, education materials, integrated alerts, and video conferencing options), *how* these elements are delivered (eg, plans and quizzes), and the subsequent intervention workflow defined as *when* they are delivered (eg, daily use)—and (2) intervention strategies—behavior change techniques [55] that underlie the intervention components (eg, *role modeling* if the Web-based education materials used include a video of patient who successfully manages his/her disease).

Outcome indicators were classified into 2 categories: effect outcome indicators and process outcome indicators [56]. Effect outcome indicators were outcomes related to self-management, health status, or cost-effectiveness, whereas process outcome indicators were outcomes on care process, health care delivery, or patient-health care interactions (eg, adherence and usability).

To allow for comparability, we classified the results reported as *positive effect*, *no statistically significant effect*, or *mixed*

effect (see Textbox 2). No negative outcomes were reported in the studies included in this review. Only quantitative methods were used to measure effect outcome indicators, whereas mixed methods were used to measure some process outcome indicators. Hence, the classification of the results of the process outcome indicators slightly differs from that of the effect outcome indicators. Outcomes related to patients and care providers are reported separately.

The determinants of implementation of eHealth self-management interventions extracted were categorized following the widely cited framework by Fleuren et al [57]. This framework identifies 50 determinants of program implementation in 5 subgroups: (1) characteristics of the sociopolitical context, such as legislation; (2) characteristics of the organization, such as staff turnover; (3) characteristics of the person adopting the innovations (user of the innovation), such as knowledge; (4) characteristics of the innovation, such as complexity; and (5) innovation strategies, such as a training. For example, the study by McGillicuddy et al [36] included in our review mentioned that "six subjects did not complete the lead-in phase, 5 for technical reasons relating to poor internet at their home." This barrier was then mirrored to the 50 determinants in Fleuren framework and classified as a determinant related to the *innovation* and, more specifically, added to the determinant category *perceived quality of eHealth intervention is excellent*. In addition, in each subgroup, we identified the influence of the patients or care providers.

Textbox 2. Outcome indicators for electronic health self-management interventions.

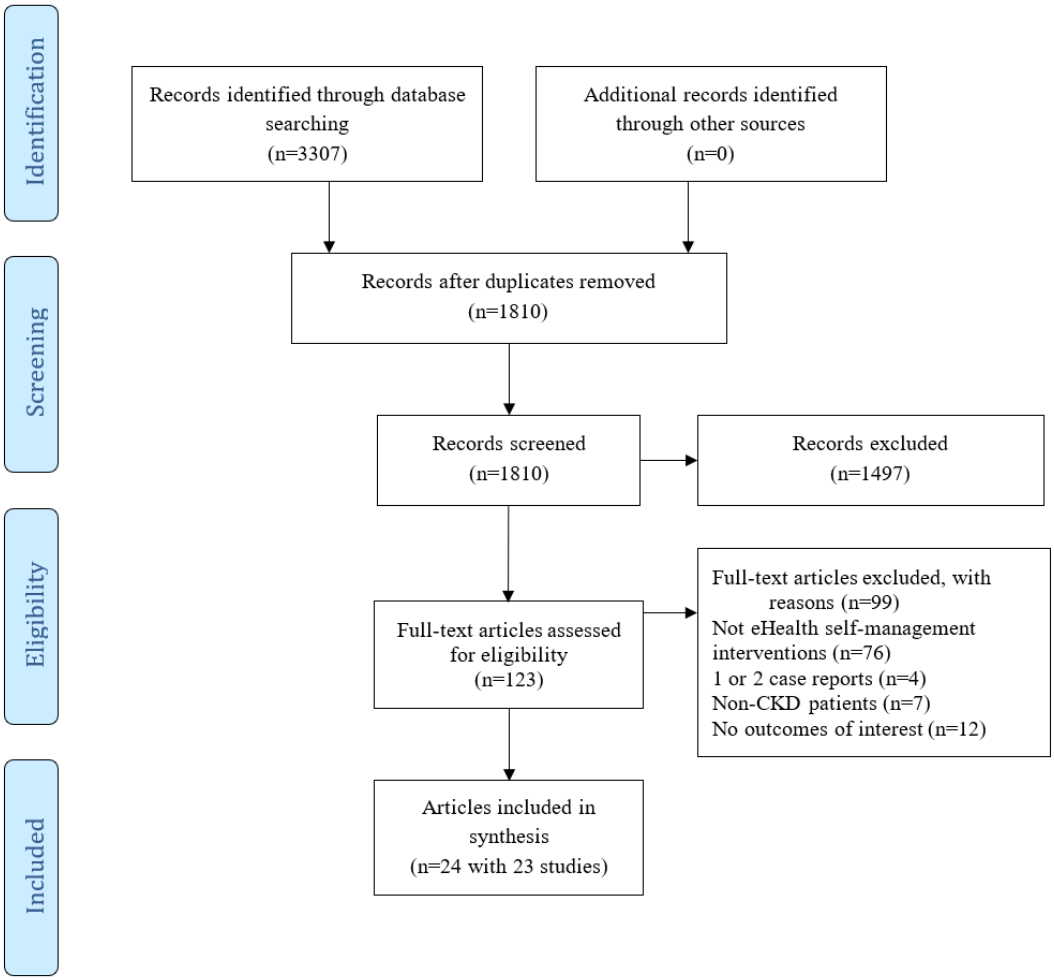
- Effect outcome indicators
  - *Positive effect*—if, after statistical analysis, a significant effect was reported
  - *No statistically significant effect*—if, after statistical analysis, a nonstatistically significant effect was reported or if no statistical analysis was performed
  - *Mixed effect*—if results that could be classified as both positive and no effect were reported
- Process outcome indicators
  - *Positive effect*—if, after statistical analysis, a statistically significant effect was reported or if a positive effect or an improvement between certain points in time was reported (eg, interviews revealed that patients were highly satisfied with the electronic health application)
  - *No statistically significant effect*—if, after statistical analysis, a nonsignificant effect was reported or if a no effect or no differences between certain points in time was reported
  - *Mixed effect*—if results that could be classified as both positive and no effect were reported

Results

Study Selection

Our search retrieved 3307 articles in total. After removing 1497 duplicates, 1810 relevant articles were screened based on title and abstract. A total of 123 potentially relevant articles were screened full text. Of these papers, 2 described results of the same RCT [58,59] and were assessed jointly. Finally, 24 articles (comprising 23 studies) [33-36,58-77] were found eligible for inclusion in this review (Figure 1).

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses flowchart of the systematic review. CKD: Chronic kidney disease; eHealth: electronic health.



Study Characteristics

All 23 studies were published between 2005 and 2017, with 19 of them being conducted between 2012 and 2017 [33-36,58,64-77]. A total of 13 studies were conducted in the United States [33-36,58,60,62-65,69,71,72], followed by 2 studies in the United Kingdom [70,74]. The research designs used varied; the majority used an RCT design [33-36,58,63,64,66,70]. Most studies focused on the usability, acceptability, and feasibility of eHealth self-management interventions [36,58,61,64,65,67,69,71,72,74-77]. Most participants are patients receiving hemodialysis [58,60,62-64,66,68,69,76,77]. Sample size at baseline ranged from 5 [67] to 601 [34]. Target population age ranged from 21 to 93 years. Recruitment mostly occurred via medical centers/hospitals [35,58,60-62,68,73]. Intervention duration ranged from 2 weeks [76] to 24 months [61]; 2 studies did not specify intervention duration [58,67]. A total of 10 studies performed a follow-up measurement [33,34,63,65,66,69-71,73,76]. Moreover, 12 studies included a control group, and 9 of those studies [33,34,36,58,61,66,68,70,76] reported *usual care* or no internet-delivered intervention as control condition. The study characteristics have been presented in Multimedia Appendix 4.

Quality Appraisal Scores of Studies

Quality of the included articles varied (Table 1). A total of 3 articles [70,73,75] were awarded a 4-star rating, 11 [33-36,59,63,65,66,69,74,76] a 3-star rating, and 10 [58,60-62,64,67,68,71,72,77] a 2-star rating or lower. Articles with a 4-star rating scored higher on design, sampling, data collection, and ethics compared with those with a 3-star rating or lower. Moreover, 20 articles [34-36,59-70,72,73,75-77] provided insufficient details on their study design or rationale. Sampling method used (eg, randomly and purposively) was not reported in 10 articles [35,60,62,65,67,71,73,74,76,77]. Although both the number and characteristics of participants were described in most articles, 15 articles [58-62,64-69,71,74,76,77] did not specify the method of sample size calculation. A total of 10 articles [58-62,64,65,67,71,72] did not detail methods used to ensure the quality of the data collected or to reduce bias. On average, the lowest score was obtained on the *ethics* section.

Table 1. Quality appraisal scores on the Crowe Critical Appraisal Tool.

Study	Total score (maximum=40)	Star score[a]	Preamble	Introduction	Design	Sample	Data	Ethic	Result	Discussion
van Lint et al (2015) [73]	33	4-star	4	4	4	4	4	4	4	5
Blakeman et al (2014) [70]	32	4-star	5	4	4	5	3	3	4	4
Ong et al (2016) [75]	32	4-star	4	4	4	4	4	4	4	4
Forni Ogna et al (2013) [66]	31	3-star	5	4	3	4	3	4	4	4
Ishani et al (2016) [34]	31	3-star	4	4	4	4	3	3	4	4
Stark et al (2011) [63]	30	3-star	4	4	4	4	3	3	4	4
McGillicuddy et al (2013) [36]	30	3-star	4	4	4	4	3	3	4	4
Hayashi et al (2017) [76]	30	3-star	5	4	3	3	4	3	4	4
Diamantidis et al (2013) [65]	29	3-star	5	4	4	3	3	4	3	3
Reese et al (2017) [35]	29	3-star	5	5	4	3	2	2	3	5
Dey et al (2016) [74]	28	3-star	3	4	3	3	3	4	4	4
Berman et al (2011) [59]	27	3-star	5	4	3	3	2	3	4	3
Rifkin et al (2013) [33]	27	3-star	5	4	3	4	3	2	3	3
Welch et al (2013) [69]	27	3-star	3	4	4	3	3	3	4	3
Connelly et al (2012) [64]	26	2-star	5	5	3	4	3	1	2	3
Neumann et al (2013) [68]	26	2-star	3	4	3	3	3	4	3	3
Liu et al (2017) [77]	25	2-star	4	5	3	2	3	1	3	4
Diamantidis et al (2015) [72]	24	2-star	3	4	3	4	1	3	3	3
Minatodani et al (2013) [58]	23	2-star	3	4	2	1	3	4	3	3
Sevick et al (2005) [60]	22	2-star	3	4	3	2	2	2	3	3
Harrington et al (2014) [71]	20	1-star	3	4	2	3	3	0	3	2
Gallar et al (2007) [61]	18	1-star	2	2	3	2	1	3	2	3
Heiden et al (2013) [67]	18	1-star	3	4	3	1	1	0	3	3
Whitten et al (2008) [62]	14	1-star	2	4	1	1	1	0	1	4

[a]1-star: more than 1 SD below mean; 2-star, between 1 SD below mean and mean; 3-star, between mean and 1 SD above mean; 4-star, more than 1 SD above mean.

Description of Electronic Health Self-Management Interventions

Major types of eHealth, functionalities, and key intervention components used are summarized in Tables 2 and 3. Most eHealth interventions evaluated included *multiple components* (multiple eHealth types) to improve patients' self-management (8/23 articles). Studies included did not provide detail on the specific intervention strategies underpinning these components, such as behavior change techniques. The most frequently used intervention component was self-monitoring (17/23 articles), followed by educational material or training (15/23 articles) and counseling (14/23 articles). Less frequently used intervention components were quizzes (3/23 articles) and interactive feedback from a device (4/23 articles). In addition, 5 studies reported that intervention development was guided by a specific theory.

Table 2. Descriptions of electronic health for each report included in the review.

Category of eHealth[a]	Detailed eHealth	Functionality
Personal digital assistant (references)		
Sevick et al (2005) [60]	Dietary self-Monitoring: meals logs	Record
Stark et al (2011) [63]	Dietary self-Monitoring: meals logs	Record
Connelly et al (2012) [64]	Dietary intake monitoring: self-monitor diet and feedback	Record
Forni Ogna et al (2013) [66]	Electronic medication event monitoring: monitor adherence	Record; communicate
Welch et al (2013) [69]	Dietary intake monitoring: self-monitor diet and feedback	Display; record
Diamantidis et al (2015) [72]	Medication inquiry system: identifying the safety of medications with impaired renal function	Record; display; alert
Telemedicine (references)		
Gallar et al (2007) [61]	Videoconferencing: connecting home to hospital	Communicate
Whitten et al (2008) [62]	Videoconferencing: connecting clinics and health system	Communicate; education
Computer (references)		
Harrington et al (2014) [71]	Tablet computer: recording data and reviewing medical findings	Display; record; communicate; alert
Ishani et al (2016) [34]	Touch screen computer with peripherals	Record; communicate
Heiden et al (2013) [67]	Educational tool, food analyzer database and diet registration, and decision support to binder dosage	Communicate; education; record
Multiple components (references)		
Diamantidis et al (2013) [65]	Alert accessories linked to website/safe kidney care: offering information	Record; education
McGillicuddy et al (2013) [36]	BP[b] monitoring, electronic medication tray, and mobile phone	Alert; communicate
Minatodani et al (2013) [58], Berman et al (2011) [59]	Self-monitoring devices	Record; communicate
Blakeman et al (2014) [70]	Website: tailoring access to community resources	Display; communicate
Dey et al (2016) [74]	Computer tablet, wearable devices, and Web portal	Record; alert
Ong et al (2016) [75]	Smartphone, a Web-based dashboard application, and a data server	Record; alert; display
Hayashi et al (2017) [76]	Self-management and recording system for dialysis (wearable devices, smartphone, and administrator module)	Record; alert; display
Liu et al (2017) [77]	App installed on mobile, cloud server, and Web app	Record; alert; communicate
Wearable devices (references)		
Neumann et al (2013) [68]	Telemetric weight monitoring	Display; alert
Rifkin et al (2013) [33]	BP monitoring	Record
van Lint et al (2015) [73]	BP monitoring and creatine monitoring	Record
Reese et al (2017) [35]	Wireless pill bottle	Record; alert

[a]eHealth: electronic health.

[b]BP: blood pressure.

Table 3. Descriptions of electronic health self-management interventions for each report included in the review.

Category of electronic health	Intervention components										Theory-based
	Educational material or training	Plan/goals	Self-monitoring	Interactive feedback from device	Message/alert to health caregivers	Message/alerts to patients from device	Message/alert to patients from health caregivers	Quizzes	Counseling	Daily use	
Personal digital assistant (references)											
Sevick et al (2005) [60]	✓	✓	✓	__[a]	—	—	✓	—	✓	✓	✓
Stark et al (2011) [63]	✓	✓	✓	—	—	—	✓	—	✓	✓	✓
Connelly et al (2012) [64]	✓	—	✓	✓	—	—	—	—	✓	—	✓
Forni Ogna et al (2013) [66]	—	✓	—	—	—	—	—	—	✓	—	—
Welch et al (2013) [69]	✓	—	✓	—	—	—	—	—	✓	—	✓
Diamantidis et al (2015) [72]	✓	—	—	✓	—	✓	—	—	—	—	—
Total (N=6), n (%)	5 (83)	3 (50)	4 (67)	2 (33)	0	1 (17)	2 (33)	0	5 (83)	2 (33)	4 (67)
Telemedicine (references)											
Gallar et al (2007) [61]	—	—	—	—	—	—	—	—	✓	—	—
Whitten et al (2008) [62]	—	—	—	—	—	—	—	—	✓	—	—
Total (N=2), n (%)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	2 (100)	0 (0)	0 (0)
Computer (references)											
Harrington et al (2014) [71]	—	✓	✓	—	✓	—	✓	—	—	✓	—
Ishani et al (2016) [34]	✓	✓	✓	—	—	—	✓	—	—	—	—
Heiden et al (2013) [67]	—	—	✓	—	—	—	—	—	—	—	—
Total (N=3), n (%)	1 (33)	2 (67)	3 (100)	0 (0)	1 (33)	0 (0)	2 (67)	0 (0)	0 (0)	1 (33)	0 (0)
Multiple components (references)											
Diamantidis et al (2013) [65]	✓	—	—	—	—	—	—	—	—	—	—
McGillicuddy et al (2013) [36]	✓	✓	✓	—	✓	✓	✓	—	✓	✓	✓

Category of electronic health	Intervention components										Theory-based
	Educational material or training	Plan/goals	Self-monitoring	Interactive feedback from device	Message/alert to health caregivers	Message/alerts to patients from device	Message/alert to patients from health caregivers	Quizzes	Counseling	Daily use	
Minatodani et al (2013) [58], Berman et al (2011) [59]	✓	✓	✓	—	—	—	✓	✓	✓	—	—
Blakeman et al (2014) [70]	✓	—	—	—	—	—	—	—	✓	—	—
Dey et al (2016) [74]	✓	—	✓	—	✓	—	✓	✓	✓	—	—
Ong et al (2016) [75]	✓	✓	✓	✓	✓	✓	—	—	—	—	—
Hayashi et al (2017) [76]	—	✓	✓	—	—	✓	✓	—	✓	✓	—
Liu et al (2017) [77]	—	—	✓	—	—	✓	—	✓	—	—	—
Total (N=8), n (%)	6 (75)	4 (50)	6 (75)	1 (13)	3 (38)	4 (50)	4 (50)	3 (38)	5 (63)	2 (25)	1 (13)
Wearable devices (references)											
Neumann et al (2013) [68]	—	✓	✓	—	✓	—	✓	—	✓	✓	—
Rifkin et al (2013) [33]	✓	—	✓	—	—	—	✓	—	✓	—	—
van Lint et al (2015) [73]	✓	✓	✓	—	—	—	—	—	—	—	—
Reese et al (2017) [35]	✓	✓	✓	✓	✓	✓	✓	—	—	—	—
Total (N=4), n (%)	3 (75)	3 (75)	4 (100)	1 (25)	2 (50)	1 (25)	3 (75)	0 (0)	2 (50)	1 (25)	0 (0)

[a]Not applicable.

Summary of Results

Tables 4 and 5 present the outcome indicators and the data collection tools used. Moreover, full details on the efficacy data reported in the included studies are included in Multimedia Appendix 5. Table 6 displays the determinants of implementation extracted. No articles reported any adverse outcomes of eHealth self-management interventions.

Table 4. Summary of outcome indicators of electronic health self-management interventions.

Outcome category and indicator	Total number of articles in each category	Effect and references		
		Positive, n (%)	No statistically significant effect, n (%)	Mixed, n (%)
Patient effect outcome (N=33)				
Blood pressure	5	4 (80) [36,68,70,75][a]	1 (20) [33]	0 (0)
Quality of life	4	1 (25) [70][a]	2 (50) [59,74]	1 (25) [76]
Laboratory tests	6	2 (33) [66,68][a]	4 (67) [60,62,75,76]	0 (0)
Interdialytic weight gain	4	1 (25) [68][a]	3 (75) [60,69,76]	0 (0)
Morbidity and mortality	2	0 (0)	2 (100) [34,61]	0 (0)
Hospitalization rate and emergency room visit	3	2 (67) [61,59][a]	1 (33) [34]	0 (0)
Medical cost	2	1 (50) [59][a]	1 (50) [61]	0 (0)
Cost-effectiveness	1	1 (100) [70][a]	0 (0)	0 (0)
Nutrition and dietary intake	2	0 (0)	2 (100) [62,69]	0 (0)
Medication adherence	4	3 (75) [35,36,66][a]	1 (25) [33]	0 (0)
Process outcome (N=28)				
Acceptability	6	6 (100); [69,74,76][b]; [33,36,75][c]	0 (0)	0 (0)
Usability	5	5 (100); [64,67,76][b]; [62,77][c]	0 (0)	0 (0)
Satisfaction	8	8 (100); [36,58,71-74,76][b]; [75][c]	0 (0)	0 (0)
Adherence to intervention	4	4(100);[35,63,73,75][b]	0 (0)	0 (0)
First entry and length of dwell time	1	1 (100); [65][b]	0 (0)	0 (0)
Self-efficacy	1	0 (0)	1 (100); [69][b]	0 (0)
Perceived benefits	1	0 (0)	1 (100); [69][b]	0 (0)
Perceived control	1	1 (100); [69][a,b]	0 (0)	0 (0)
Recorded errors	1	1 (100); [72][b]	0 (0)	0 (0)

[a]Statistically significant.

[b]Outcome related to patient.

[c]Outcome related to both patient and care provider.

Table 5. Summary of reported tools of outcome indicators.

Outcome category and indicator	Reported data collection tools (number of articles)
Patient effect outcome (N=33), all quantitative	
Blood pressure	Readings (4) and dataset (1)
Quality of life	36-item Short Form Health Survey (1), EuroQoL-5 Dimension (1), and 36-item Kidney Disease Quality of Life survey (2)
Laboratory tests	Medical records (6)
Interdialytic weight gain	Medical records (4)
Morbidity and mortality	Charlson comorbidity index (1) and records (1)
Hospitalization rate and emergency room visit	Records (3)
Medical cost	Records (2)
Cost-effectiveness	Records (1)
Nutrition and dietary intake	Clinical data (2)
Medication adherence	System data (2), adherence score calculation (1), and Morisky Medication Adherence Scale (1)
Process outcome (N=28)	
Acceptability	Quantitative: questionnaires (1), recruitments and participation rate (1), QUEST[a] and retention rates (1), and average number of daily entries and completion rates (2); quantitative and qualitative: number of assessments and semistructured interview (1)
Usability	Quantitative: survey (1) and questionnaire (2); qualitative: interview (1); quantitative and qualitative: survey, interview, and system data (1)
Satisfaction	Quantitative: questionnaires and QUEST (5); qualitative: semistructured interview (2); quantitative and qualitative: questionnaire and interview (1)
Adherence to intervention	Quantitative: system data (3) and Basel Assessment of Adherence to Immunosuppressive Medications Scale (1)
First entry and length of dwell time	Quantitative: frequency and number (1)
Self-efficacy	Quantitative: cardiac diet self-efficacy and Fluid Self-Efficacy Scale (1)
Perceived benefits	Quantitative: Benefits of Sodium Adherence and a 9-item Benefits of Fluid Adherence Scale (1)
Perceived control	Quantitative: 7-item Mastery scale (1)
Recorded errors	Quantitative: questionnaire and record (1)

[a]QUEST: Quebec user evaluation of satisfaction with assistive technology.

Table 6. Determinants of the implementation of electronic health self-management interventions for chronic kidney disease.

Determinants of interventions and details	References	
	If determinant is present	If determinant is exact opposite
Sociopolitical context (patient)		
Awareness of potential health benefits of the eHealth[a] self-management intervention	Berman et al (2011) [59], Hayashi et al (2017) [76]	—[b]
Target population feels comfortable about eHealth use	Hayashi et al (2017) [76], Liu et al (2017) [77]	—
Organization (patient)		
Community resources (eg, activities, services, and applicable wireless fidelity connection at the users' location) available for implementation	Blakeman et al (2014) [70]	Harrington et al (2014) [71]
User		
Patient		
Support from colleagues (eg, internet personnel)	Stark et al (2011) [63]	McGillicuddy et al (2013) [36]
Ability of health care professionals to monitor and, if necessary, anticipate on patient measurements online	Rifkin et al (2013) [33], Reese et al (2017) [35], Berman et al (2011) [59], van Lint et al (2015) [73], Ong et al (2016) [75], Liu et al (2017) [77]	—
Availability of sufficient skills/knowledge	Diamantidis et al (2015) [72]	Berman et al (2011) [59], Welch et al (2013) [69]
eHealth technology is considered valuable by user	McGillicuddy et al (2013) [36], Heiden et al (2013) [67], Blakeman et al (2014) [70]	van Lint et al (2015) [73]
High self-efficacy	van Lint et al (2015) [73]	—
Patient and care provider		
eHealth technology is considered valuable by user	Rifkin et al (2013) [33]	—
Innovation		
Patient		
Implementation of intervention is perceived as risk-free by user	Harrington et al (2014) [71], Dey et al (2016) [74], Hayashi et al (2017) [76]	—
Provision of warning/alert/reminder based on parameters monitored	Reese et al (2017) [35], McGillicuddy et al (2013) [36], van Lint et al (2015) [73]	—
Provision of real-time feedback (eg, amount of dietary intake, blood pressure value) based on patients' input	Sevick et al (2005) [60], Stark et al (2011) [63], Connelly et al (2012) [64], Hayashi et al (2017) [76]	—
Perceived quality of eHealth intervention is excellent	—	Rifkin et al (2013) [33], McGillicuddy et al (2013) [36], Berman et al (2011) [59], Harrington et al (2014) [71]
Patient and care provider		
Interventions are compatible with existing work procedures	Rifkin et al (2013) [33]	—
Implementation of intervention is perceived as advantageous by patient and care providers considering increasing access to health care services	Whitten et al (2008) [62]	—
High acceptability of eHealth	Rifkin et al (2013) [33]	—
Perceived quality of eHealth intervention is excellent	Gallar et al (2007) [61]	—
Innovation strategies (patient and care provider)		
Well planned/structured implementation process	Liu et al (2017) [77]	—

[a]eHealth: electronic health.

[b]Not applicable.

Description of Effect Outcome Indicators

The effect outcome indicators most frequently reported were laboratory tests (eg, serum albumin, C-reactive protein; 6/23 articles) and blood pressure (BP; 5/23 articles). Interdialytic weight gain (4/23 articles), QoL (4/23 articles), and medication adherence (4/23 articles) were also frequently reported. Finally, 2 studies assessed effects on morbidity and mortality, 2 evaluated changes in medical cost, and 1 performed a cost-effectiveness analysis.

Out of 5 studies, 4 [36,68,70,75] reported a statistically significant positive effect on BP. Of the 2 studies [59,61] that evaluated changes in medical costs, 1 [59] reported a significant reduction in costs in the intervention group. A study reported an incremental cost-effectiveness ratio of US $175, showing that the implementation of a website-based self-management intervention for CKD patients was superior, considering effects and costs, to usual care [70]. Out of 3 studies, 2 [59,61] reported statistically significant improvements in hospitalization rates and emergency room visits. Out of 4 studies, 3 [35,36,66] reported statistically significant improvements in patients' medication adherence. Out of 4 studies, 1 [70] reported a statistically significant improvement on QoL.

Description of Process Outcome Indicators

The process outcome indicator *satisfaction* was reported in one-third of included studies. A total of 2 studies [58,75] used interviews to evaluate satisfaction in patients or care providers. Patients were reported to be satisfied with the use of at-home telehealth and appreciated its utility in managing their health [58]. Patients using a smartphone-based self-management system indicated feeling more confident and more in control of their condition; the nurses found that the system helped prioritize patients who needed more attention [75]. A total of 5 studies used questionnaires to evaluate satisfaction of patients [36,71,72,74,76]. These studies reported patients were highly satisfied with eHealth self-management interventions.

Acceptability was also frequently reported and mostly measured using questionnaires, retention rates, or system data [33,36,69,74-76] (6/23 articles). All these studies reported that eHealth self-management interventions were acceptable to patients [33,36,69,74-76] and care providers [33,36]. Other process outcome indicators (including adherence to the intervention, first entry, length of dwell time, self-efficacy, perceived benefits, perceived control, and recorded errors) were less frequently used.

Description of Implementation Determinants

All but 4 studies [34,65,66,68] reported on determinants of implementation. Studies included used various methods (eg, qualitative interview and quantitative data analysis) to evaluate determinants of implementation. The determinant *ability of health care professionals to monitor and, if necessary, anticipate on patient measurements online* is mostly reported to make patients feel safe while using eHealth interventions [77], thereby influencing patients' medication adherence [35] and adherence to interventions [35,73]. Moreover, *availability of sufficient skills/knowledge* [58,69,72] was reported as an important determinant to patients' use of the eHealth self-management interventions. In addition, the determinant *provision of real-time feedback based on patients' input* was frequently reported to influence patients' adherence to self-monitoring and healthy behaviors [60,63,64,76]. The determinant *perceived quality of eHealth intervention is excellent* [61] was cited to influence both patients' and care providers' use of the intervention. The percent agreement between the 2 reviewers' classification of the implementation determinants reported following the Fleuren framework was 76%, which is considered acceptable [48]. Discrepancies in classification were discussed until consensus was reached.

Discussion

Principal Findings

The main findings and implications have been presented in Textbox 3.

The evidence regarding the implementation and effectiveness of eHealth self-management interventions for CKD patients was reviewed. The 23 studies included were appraised on methodological quality, and all relevant data were extracted. Although the evidence base is still inconclusive, our review provides an indication that eHealth self-management interventions have the potential to improve CKD patients' management and health outcomes. Furthermore, high acceptability of and satisfaction with the eHealth interventions used were reported. Owing to the heterogeneity of the intervention components and outcomes measures used, we could not determine which intervention components contributed most to the effects found. The determinant *ability of health care professionals to monitor and, if necessary, anticipate on patient measurements online* was most frequently reported to influence implementation. The determinants reported were not quantified, and the relative importance of each determinant could not be determined.

Textbox 3. Main findings and implications for this study.

- Although the evidence base is still inconclusive, a majority of studies on electronic health (eHealth) self-management interventions report improvements on proximal outcomes (eg, blood pressure controlling) and mixed effects for more distal (eg, quality of life) outcomes.
- Evidence on the process level is more established; eHealth self-management interventions for chronic kidney disease patients are reported to be highly feasible, usable, and acceptable.
- To adequately assess eHealth intervention effect, future researchers should carefully consider their choice of outcomes (distal vs proximal) based on their sensitivity to capture meaningful change.
- Standardization of research design and methods in the evaluation of eHealth self-management interventions for chronic kidney disease patients is needed to optimize quality and comparability across studies and further elucidate which intervention components alone or in interaction contribute to the promising results found.

Comparison of Findings

Most studies reported the evaluation of effect outcome indicators. The positive effects on patients' BP controlling [36,68,70,75] and medication adherence [35,36,66] were consistently reported; no adverse outcomes were reported. These findings correspond with another review on eHealth interventions in CKD [39]. Compared with standard outpatient-based management, eHealth self-management interventions have the potential to reduce health care delivery costs [78]. Although this potential reduction in costs is essential for policy makers and clinicians to adopt eHealth self-management interventions, health care expenditures were only assessed in 3 of the studies included, with only 1 performing a cost-effectiveness analysis [70]. Hence, we cannot yet determine if and how these interventions might reduce medical costs. This finding is consistent with similar reviews, which conclude that studies on the cost-effectiveness of eHealth self-management interventions are either conflicting or lacking [32,54]. As evidence on cost-effectiveness is important to support the potential scale-up of eHealth technology, further research is needed to broaden this evidence base. Regarding QoL, only 1 out of 4 studies reported a significant improvement. A possible explanation for this finding was the short follow-up period instated to capture changes in a distal outcome such as QoL [59]. As QoL in CKD is an independent predictor of mortality and hospitalization [79,80], and thus important to evaluate, we advise further research to assess QoL with a longer follow-up period.

In general, we found that eHealth self-management interventions were reported to be highly feasible, usable, and acceptable. However, we found great diversity in the use and operationalization of outcome indicators and how they were measured. For instance, a study reported acceptability by measuring adoption, adherence to the recommended intervention use, user satisfaction, and feature usage [75]. In contrast, other studies [33,36] measured acceptability by asking patients "how acceptable they found the intervention" using a self-report scale. It is also notable that only 4 studies assessed implementation adherence, although finding no or limited intervention effects can be strongly related to patients' nonadherence to eHealth interventions as prescribed [81,82]. Examining implementation adherence can help resolve the *black box* of patients' adoption and continued use of the intervention, thereby preventing a type 3 error [83]. To tackle these issues, we advise researchers to use a standardized operationalization of process outcome indicators and measure implementation adherence to enable reliable interpretation of the intervention effect found.

Considering which outcomes are most sensitive to change is important. As eHealth interventions studies are mostly of short duration, they may not detect changes in distal outcomes (eg, QoL). Hence, effectivity might be easier to detect when proximal outcomes, close to the intervention strategies, are measured. For example, BP controlling can be an outcome sensitive to change if self-monitoring is the main intervention component. Functional outcomes (such as days needed to return to work), which can quantify patients' subjective perceptions of the effect of treatment on their daily life, might also be very sensitive to change by eHealth interventions [84,85]. Moreover, researchers should consider if their outcomes reflect meaningful change and provide a clear rationale for their choice of laboratory parameters. For example, using serum albumin as an indicator for dietary adherence might be of limited value as it is influenced by other CKD characteristics (eg, low dialysis dose) [60].

Furthermore, improving knowledge on the effect modifiers at play in eHealth self-management interventions for CKD patients is important. None of the included studies provided detail on potentially relevant effect modifiers. We can identify some possible modifying factors based on research focusing on self-management interventions in other chronic, noncommunicable diseases (NCDs). For instance, a longer intervention duration might positively modify the effect of self-management interventions [86]. In addition, the patients' health literacy level might modify intervention effect [87]. Self-management interventions for NCDs are mostly based on similar intervention principles and behavior change techniques. Moreover, the characteristics of patients suffering from NCDs are often similar. We, therefore, argue that the modifiers found to influence the outcomes of self-management interventions for NCDs in general might also be applicable for similar interventions targeting CKD patients. However, more research is needed to identify effect modifiers to self-management interventions targeting CKD and explore possible strategies to impact these factors.

Electronic Health Self-Management Interventions

A large variety of eHealth self-management intervention components were used in the included studies (eg, self-monitoring and education), and the results differed greatly. These findings make it difficult and possibly premature to formulate a potentially ideal palette of eHealth self-management intervention components for CKD patients. However, reviewing results make it possible to identify which intervention components might be more promising than others. For instance, self-monitoring and the use of messages or alerts to nudge patient toward displaying healthy behaviors (see Multimedia Appendix 6) were most commonly reported as the effective components to optimize patient self-management skills.

Furthermore, few of the interventions studied were theory-based. The authors recommend that a strong theoretical foundation is necessary for the planning, design, evaluation, and implementation of eHealth self-management interventions [88]. We recommend building eHealth self-management interventions based on established behavior change techniques, such as formulated in the Behavior Change Techniques taxonomy [55]. Moreover, the use of cocreation methods and appreciative inquiry (such as described in the Center for eHealth Research and Disease Management [89] roadmap for eHealth development) can improve intervention fit with the needs and priorities expressed by professionals and patients.

Determinants of Implementation

Ability of health care professionals to monitor and, if necessary, anticipate patient measurements online was reported as an important determinant of implementation. We argue that this ability of professionals to anticipate and act upon patient measurements might reduce patients' feeling of isolation and/or

anxiety caused by independently conducted treatments at home [77] and thereby increase patients' adherence to implementation. In addition, *availability of sufficient skills/knowledge* was important for users to continue their use of eHealth technology. If participants are unfamiliar with the use of eHealth, this has been reported to limit their acceptance of eHealth interventions [58,69]. Proper training and tailored tutorials are needed to guide eHealth implementation to optimize knowledge and skills and promote intervention uptake [67,72]. The included studies used various methods to evaluate determinants of implementation. We suggest that future research should use validated tools for measuring implementation quality and related determinants, such as the Measurement Instrument for Determinants of Innovations questionnaire and Determinants of Implementation Behavior Questionnaire [90,91].

Study Quality and Characteristics

Most studies were appraised to be of low to moderate quality. There is a heterogeneity of outcome measurement tools and reporting styles used in the articles included in this review. Therefore, we advise researchers to develop a more standardized approach to the use of outcome measures, guided by, for instance, the formulation of an International Consortium for Health Outcomes Measurement standard set for CKD [92]. In addition, we argue that detailed description and a thorough analysis of study design, methods, and intervention components used, based on a published theoretical framework such as Consolidated Standards of Reporting Trials-eHealth [93], can improve reporting and provide a basis for evaluating the validity and applicability of eHealth trials.

Data on eHealth self-management interventions for CKD patients in developing countries are still lacking, which corresponds with other reviews on eHealth interventions [94,95]. The need to perform such research in developing countries is high. eHealth interventions in these countries have the potential to improve the accessibility and cost-effectiveness of local care and ensure timely delivery of care to rural areas and diverse populations [20,24,96]. Furthermore, 9 studies had an intervention duration of fewer than 6 months. Few studies conducted a follow-up measurement. Forni Ogna et al [66] reported that the positive intervention effects were maintained only during the monitoring period; these effects had vanished 3 months after interruption of the drug adherence monitoring. This finding underlines that the effectiveness of eHealth self-management interventions should be tested during a longer study period and with follow-up measurements.

Of note, 3 studies with fewer than 10 participants were included. One might argue that such studies do not provide robust, generalizable evidence and should be excluded based only on their sample size. However, high-level evidence on the effectiveness of eHealth self-management interventions for CKD patients, for instance, generated by large RCTs, is very limited. Hence, studies with less robust designs are included, as in this stage, we feel that all evidence should be accumulated and taken into account as to broaden our view and deepen our understanding of the usability, implementability, and effectiveness of eHealth self-management interventions for CKD patients. Moreover, this decision is supported by similar systematic reviews on the effectivity of eHealth interventions that also included studies with smaller sample sizes [95,97,98]. That being said, results of this review should be interpreted with some caution.

Strengths and Limitations

To our knowledge, this is the first systematic review to evaluate the entire spectrum of studies focusing on eHealth self-management interventions for CKD patients. Our review has some strengths. First, PRISMA guidelines were followed, and a robust search strategy was used in 8 databases. Second, a comprehensive analysis was conducted on the intervention components, outcome indicators, and determinants from the various studies. The kappa value and percent agreement obtained, and thus inter-rater reliability, showed that the validity of the appraisal could be considered fair. Finally, any discrepancies were discussed until consensus was reached.

Nevertheless, several limitations need to be addressed. First, as articles only published in English were included, some relevant articles might have been missed. Second, substantial heterogeneity of interventions and outcome measures made it difficult to draw firm conclusions about the evidence emerging from these studies, and results should be interpreted with caution.

Conclusions

This review provides a comprehensive overview of studies evaluating eHealth self-management interventions for CKD patients. eHealth self-management interventions show promise to improve health outcomes in CKD patients. To adequately assess eHealth intervention effect, future researchers should carefully consider their choice of outcomes (distal vs proximal) based on their sensitivity to capture meaningful change. Also, to enable the standard design and scale-up of effective eHealth self-management interventions for CKD patients, a more detailed understanding of which individual intervention components lead to health outcome improvement and which determinants of the implementation can promote adherence and satisfaction with care is needed.

Acknowledgments

This study was funded by the China Scholarship Council.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[PDF File (Adobe PDF File), 222 KB - jmir_v21i11e12384_app1.pdf]

Multimedia Appendix 2
Major types of electronic health.
[PDF File (Adobe PDF File), 193 KB - jmir_v21i11e12384_app2.pdf]

Multimedia Appendix 3
Electronic health self-management intervention components.
[PDF File (Adobe PDF File), 283 KB - jmir_v21i11e12384_app3.pdf]

Multimedia Appendix 4
Study characteristics.
[PDF File (Adobe PDF File), 170 KB - jmir_v21i11e12384_app4.pdf]

Multimedia Appendix 5
Effects and references of outcome indicators.
[PDF File (Adobe PDF File), 3768 KB - jmir_v21i11e12384_app5.pdf]

Multimedia Appendix 6
Most frequently recommended electronic health self-management intervention components.
[PDF File (Adobe PDF File), 193 KB - jmir_v21i11e12384_app6.pdf]

References

1. Hill NR, Fatoba ST, Oke JL, Hirst JA, O'Callaghan CA, Lasserson DS, et al. Global prevalence of chronic kidney disease - a systematic review and meta-analysis. PLoS One 2016;11(7):e0158765 [FREE Full text] [doi: 10.1371/journal.pone.0158765] [Medline: 27383068]
2. Zhang L, Wang F, Wang L, Wang W, Liu B, Liu J, et al. Prevalence of chronic kidney disease in China: a cross-sectional survey. Lancet 2012 Mar 3;379(9818):815-822. [doi: 10.1016/S0140-6736(12)60033-6] [Medline: 22386035]
3. GBD 2013 Mortality and Causes of Death Collaborators. Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990-2013: a systematic analysis for the global burden of disease study 2013. Lancet 2015 Jan 10;385(9963):117-171 [FREE Full text] [doi: 10.1016/S0140-6736(14)61682-2] [Medline: 25530442]
4. Zhang Q, Rothenbacher D. Prevalence of chronic kidney disease in population-based studies: systematic review. BMC Public Health 2008 Apr 11;8:117 [FREE Full text] [doi: 10.1186/1471-2458-8-117] [Medline: 18405348]
5. National Kidney Foundation. K/DOQI clinical practice guidelines for chronic kidney disease: evaluation, classification, and stratification. Am J Kidney Dis 2002 Feb;39(2 Suppl 1):S1-266. [Medline: 11904577]
6. Couser WG, Remuzzi G, Mendis S, Tonelli M. The contribution of chronic kidney disease to the global burden of major noncommunicable diseases. Kidney Int 2011 Dec;80(12):1258-1270 [FREE Full text] [doi: 10.1038/ki.2011.368] [Medline: 21993585]
7. Gansevoort RT, Correa-Rotter R, Hemmelgarn BR, Jafar TH, Heerspink HJL, Mann JF, et al. Chronic kidney disease and cardiovascular risk: epidemiology, mechanisms, and prevention. Lancet 2013 Jul 27;382(9889):339-352. [doi: 10.1016/S0140-6736(13)60595-4] [Medline: 23727170]
8. Braun L, Sood V, Hogue S, Lieberman B, Copley-Merriman C. High burden and unmet patient needs in chronic kidney disease. Int J Nephrol Renovasc Dis 2012;5:151-163 [FREE Full text] [doi: 10.2147/IJNRD.S37766] [Medline: 23293534]
9. Chin HJ, Song YR, Lee JJ, Lee SB, Kim KW, Na KY, et al. Moderately decreased renal function negatively affects the health-related quality of life among the elderly Korean population: a population-based study. Nephrol Dial Transplant 2008 Sep;23(9):2810-2817. [doi: 10.1093/ndt/gfn132] [Medline: 18372390]
10. Etgen T, Chonchol M, Förstl H, Sander D. Chronic kidney disease and cognitive impairment: a systematic review and meta-analysis. Am J Nephrol 2012;35(5):474-482 [FREE Full text] [doi: 10.1159/000338135] [Medline: 22555151]
11. Lorig KR, Holman HR. Self-management education: history, definition, outcomes, and mechanisms. Ann Behav Med 2003 Aug;26(1):1-7. [doi: 10.1207/S15324796ABM2601_01] [Medline: 12867348]
12. Bonner A, Havas K, Douglas C, Thepha T, Bennett P, Clark R. Self-management programmes in stages 1-4 chronic kidney disease: a literature review. J Ren Care 2014 Sep;40(3):194-204. [doi: 10.1111/jorc.12058] [Medline: 24628848]
13. Moattari M, Ebrahimi M, Sharifi N, Rouzbeh J. The effect of empowerment on the self-efficacy, quality of life and clinical and laboratory indicators of patients treated with hemodialysis: a randomized controlled trial. Health Qual Life Outcomes 2012 Sep 20;10:115 [FREE Full text] [doi: 10.1186/1477-7525-10-115] [Medline: 22992449]

14. Boots LM, de Vugt ME, Kempen GI, Verhey FR. Effectiveness of a blended care self-management program for caregivers of people with early-stage dementia (Partner In Balance): randomized controlled trial. J Med Internet Res 2018 Jul 13;20(7):e10017 [FREE Full text] [doi: 10.2196/10017] [Medline: 30006327]
15. Karamanidou C, Clatworthy J, Weinman J, Horne R. A systematic review of the prevalence and determinants of nonadherence to phosphate binding medication in patients with end-stage renal disease. BMC Nephrol 2008 Jan 31;9:2 [FREE Full text] [doi: 10.1186/1471-2369-9-2] [Medline: 18237373]
16. Ghimire S, Castelino RL, Lioufas NM, Peterson GM, Zaidi ST. Nonadherence to medication therapy in haemodialysis patients: a systematic review. PLoS One 2015;10(12):e0144119 [FREE Full text] [doi: 10.1371/journal.pone.0144119] [Medline: 26636968]
17. Donald M, Kahlon BK, Beanlands H, Straus S, Ronksley P, Herrington G, et al. Self-management interventions for adults with chronic kidney disease: a scoping review. BMJ Open 2018 Mar 22;8(3):e019814 [FREE Full text] [doi: 10.1136/bmjopen-2017-019814] [Medline: 29567848]
18. Havas K, Douglas C, Bonner A. Closing the loop in person-centered care: patient experiences of a chronic kidney disease self-management intervention. Patient Prefer Adherence 2017;11:1963-1973 [FREE Full text] [doi: 10.2147/PPA.S147831] [Medline: 29238172]
19. Slater H, Dear BF, Merolli MA, Li LC, Briggs AM. Use of ehealth technologies to enable the implementation of musculoskeletal models of care: evidence and practice. Best Pract Res Clin Rheumatol 2016 Jun;30(3):483-502 [FREE Full text] [doi: 10.1016/j.berh.2016.08.006] [Medline: 27886943]
20. Schwamm LH, Chumbler N, Brown E, Fonarow GC, Berube D, Nystrom K, American Heart Association Advocacy Coordinating Committee. Recommendations for the implementation of telehealth in cardiovascular and stroke care: a policy statement from the American Heart Association. Circulation 2017 Feb 14;135(7):e24-e44. [doi: 10.1161/CIR.0000000000000475] [Medline: 27998940]
21. Larsen-Cooper E, Bancroft E, Rajagopal S, O'Toole M, Levin A. Scale matters: a cost-outcome analysis of an m-health intervention in Malawi. Telemed J E Health 2016 Apr;22(4):317-324 [FREE Full text] [doi: 10.1089/tmj.2015.0060] [Medline: 26348994]
22. Eysenbach G. What is e-health? J Med Internet Res 2001;3(2):E20 [FREE Full text] [doi: 10.2196/jmir.3.2.e20] [Medline: 11720962]
23. Slev VN, Pasman HR, Eeltink CM, van Uden-Kraan CF, Verdonck-de Leeuw IM, Francke AL. Self-management support and ehealth for patients and informal caregivers confronted with advanced cancer: an online focus group study among nurses. BMC Palliat Care 2017 Nov 21;16(1):55 [FREE Full text] [doi: 10.1186/s12904-017-0238-4] [Medline: 29162081]
24. Black AD, Car J, Pagliari C, Anandan C, Cresswell K, Bokun T, et al. The impact of ehealth on the quality and safety of health care: a systematic overview. PLoS Med 2011 Jan 18;8(1):e1000387 [FREE Full text] [doi: 10.1371/journal.pmed.1000387] [Medline: 21267058]
25. Bertz F, Winkvist A, Brekke HK. Sustainable weight loss among overweight and obese lactating women is achieved with an energy-reduced diet in line with dietary recommendations: results from the LEVA randomized controlled trial. J Acad Nutr Diet 2015 Jan;115(1):78-86. [doi: 10.1016/j.jand.2014.05.017] [Medline: 25088520]
26. Herring SJ, Cruice JF, Bennett GG, Rose MZ, Davey A, Foster GD. Preventing excessive gestational weight gain among African American women: a randomized clinical trial. Obesity (Silver Spring) 2016 Jan;24(1):30-36 [FREE Full text] [doi: 10.1002/oby.21240] [Medline: 26592857]
27. Taylor GM, Dalili MN, Semwal M, Civljak M, Sheikh A, Car J. Internet-based interventions for smoking cessation. Cochrane Database Syst Rev 2017 Sep 4;9:CD007078 [FREE Full text] [doi: 10.1002/14651858.CD007078.pub5] [Medline: 28869775]
28. Clarke G, Eubanks D, Reid E, Kelleher C, O'Connor E, DeBar LL, et al. Overcoming depression on the internet (ODIN) (2): a randomized trial of a self-help depression skills program with reminders. J Med Internet Res 2005 Jun;7(2):e16 [FREE Full text] [doi: 10.2196/jmir.7.2.e16] [Medline: 15998607]
29. Steventon A, Bardsley M, Billings J, Dixon J, Doll H, Hirani S, Whole System Demonstrator Evaluation Team. Effect of telehealth on use of secondary care and mortality: findings from the Whole System Demonstrator cluster randomised trial. Br Med J 2012 Jun 21;344:e3874 [FREE Full text] [doi: 10.1136/bmj.e3874] [Medline: 22723612]
30. Seiler A, Klaas V, Tröster G, Fagundes CP. eHealth and mHealth interventions in the treatment of fatigued cancer survivors: a systematic review and meta-analysis. Psychooncology 2017 Sep;26(9):1239-1253. [doi: 10.1002/pon.4489] [Medline: 28665554]
31. Palacios J, Lee GA, Duaso M, Clifton A, Norman IJ, Richards D, et al. Internet-delivered self-management support for improving coronary heart disease and self-management-related outcomes: a systematic review. J Cardiovasc Nurs 2017;32(4):E9-23 [FREE Full text] [doi: 10.1097/JCN.0000000000000392] [Medline: 28107251]
32. Jackson BD, Gray K, Knowles SR, de Cruz P. Ehealth technologies in inflammatory bowel disease: a systematic review. J Crohns Colitis 2016 Sep;10(9):1103-1121. [doi: 10.1093/ecco-jcc/jjw059] [Medline: 26928960]
33. Rifkin DE, Abdelmalek JA, Miracle CM, Low C, Barsotti R, Rios P, et al. Linking clinic and home: a randomized, controlled clinical effectiveness trial of real-time, wireless blood pressure monitoring for older patients with kidney disease and hypertension. Blood Press Monit 2013 Feb;18(1):8-15 [FREE Full text] [doi: 10.1097/MBP.0b013e32835d126c] [Medline: 23275313]

34. Ishani A, Christopher J, Palmer D, Otterness S, Clothier B, Nugent S, Center for Innovative Kidney Care. Telehealth by an interprofessional team in patients with CKD: a randomized controlled trial. Am J Kidney Dis 2016 Jul;68(1):41-49. [doi: 10.1053/j.ajkd.2016.01.018] [Medline: 26947216]
35. Reese PP, Bloom RD, Trofe-Clark J, Mussell A, Leidy D, Levsky S, et al. Automated reminders and physician notification to promote immunosuppression adherence among kidney transplant recipients: a randomized trial. Am J Kidney Dis 2017 Mar;69(3):400-409. [doi: 10.1053/j.ajkd.2016.10.017] [Medline: 27940063]
36. McGillicuddy JW, Gregoski MJ, Weiland AK, Rock RA, Brunner-Jackson BM, Patel SK, et al. Mobile health medication adherence and blood pressure control in renal transplant recipients: a proof-of-concept randomized controlled trial. JMIR Res Protoc 2013 Sep 4;2(2):e32 [FREE Full text] [doi: 10.2196/resprot.2633] [Medline: 24004517]
37. Campbell J, Porter J. Dietary mobile apps and their effect on nutritional indicators in chronic renal disease: a systematic review. Nephrology (Carlton) 2015 Oct;20(10):744-751. [doi: 10.1111/nep.12500] [Medline: 25959301]
38. He T, Liu X, Li Y, Wu Q, Liu M, Yuan H. Remote home management for chronic kidney disease: a systematic review. J Telemed Telecare 2017 Jan;23(1):3-13. [doi: 10.1177/1357633X15626855] [Medline: 27269795]
39. Jeddi FR, Nabovati E, Amirazodi S. Features and effects of information technology-based interventions to improve self-management in chronic kidney disease patients: a systematic review of the literature. J Med Syst 2017 Sep 18;41(11):170. [doi: 10.1007/s10916-017-0820-6] [Medline: 28920174]
40. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
41. Kaltenthaler E, Sutcliffe P, Parry G, Beverley C, Rees A, Ferriter M. The acceptability to patients of computerized cognitive behaviour therapy for depression: a systematic review. Psychol Med 2008 Nov;38(11):1521-1530. [doi: 10.1017/S0033291707002607] [Medline: 18205964]
42. Crowe M, Sheppard L. A general critical appraisal tool: an evaluation of construct validity. Int J Nurs Stud 2011 Dec;48(12):1505-1516. [doi: 10.1016/j.ijnurstu.2011.06.004] [Medline: 21802686]
43. Crowe M, Sheppard L, Campbell A. Comparison of the effects of using the Crowe Critical Appraisal Tool versus informal appraisal in assessing health research: a randomised trial. Int J Evid Based Healthc 2011 Dec;9(4):444-449. [doi: 10.1111/j.1744-1609.2011.00237.x] [Medline: 22093394]
44. Crowe M, Sheppard L, Campbell A. Reliability analysis for a proposed critical appraisal tool demonstrated value for diverse research designs. J Clin Epidemiol 2012 Apr;65(4):375-383. [doi: 10.1016/j.jclinepi.2011.08.006] [Medline: 22078576]
45. Corrigan FM, Broome H, Dorris L. A systematic review of psychosocial interventions for children and young people with epilepsy. Epilepsy Behav 2016 Mar;56:99-112. [doi: 10.1016/j.yebeh.2016.01.005] [Medline: 26859319]
46. Crowe L, Chang A, Fraser JA, Gaskill D, Nash R, Wallace K. Systematic review of the effectiveness of nursing interventions in reducing or relieving post-operative pain. Int J Evid Based Healthc 2008 Dec;6(4):396-430. [doi: 10.1111/j.1744-1609.2008.00113.x] [Medline: 21631835]
47. van der Kleij R, Coster N, Verbiest M, van Assema P, Paulussen T, Reis R, et al. Implementation of intersectoral community approaches targeting childhood obesity: a systematic review. Obes Rev 2015 Jun;16(6):454-472. [doi: 10.1111/obr.12273] [Medline: 25824957]
48. Fleiss JL. Statistical Methods for Rates and Proportions. Second Edition. New York: Wiley-Interscience; 1981.
49. Lipsey MW, Wilson DB. Practical Meta-Analysis. First Edition. Thousand Oaks, CA: Sage Publications; 2001.
50. Vegesna A, Tran M, Angelaccio M, Arcona S. Remote patient monitoring via non-invasive digital technologies: a systematic review. Telemed J E Health 2017 Jan;23(1):3-17 [FREE Full text] [doi: 10.1089/tmj.2016.0051] [Medline: 27116181]
51. Powell AC, Landman AB, Bates DW. In search of a few good apps. J Am Med Assoc 2014 May 14;311(18):1851-1852. [doi: 10.1001/jama.2014.2564] [Medline: 24664278]
52. Chaet AV, Morshedi B, Wells KJ, Barnes LE, Valdez R. Spanish-language consumer health information technology interventions: a systematic review. J Med Internet Res 2016 Aug 10;18(8):e214 [FREE Full text] [doi: 10.2196/jmir.5794] [Medline: 27511437]
53. Mohr DC, Schueller SM, Riley WT, Brown CH, Cuijpers P, Duan N, et al. Trials of intervention principles: evaluation methods for evolving behavioral intervention technologies. J Med Internet Res 2015 Jul 8;17(7):e166 [FREE Full text] [doi: 10.2196/jmir.4391] [Medline: 26155878]
54. Morrison D, Wyke S, Agur K, Cameron EJ, Docking RI, Mackenzie AM, et al. Digital asthma self-management interventions: a systematic review. J Med Internet Res 2014 Feb 18;16(2):e51 [FREE Full text] [doi: 10.2196/jmir.2814] [Medline: 24550161]
55. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
56. Beratarrechea A, Lee AG, Willner JM, Jahangir E, Ciapponi A, Rubinstein A. The impact of mobile health interventions on chronic disease outcomes in developing countries: a systematic review. Telemed J E Health 2014 Jan;20(1):75-82 [FREE Full text] [doi: 10.1089/tmj.2012.0328] [Medline: 24205809]

57. Fleuren M, Wiefferink K, Paulussen T. Determinants of innovation within health care organizations: literature review and Delphi study. Int J Qual Health Care 2004 Apr;16(2):107-123. [doi: 10.1093/intqhc/mzh030] [Medline: 15051705]
58. Minatodani DE, Chao PJ, Berman SJ. Home telehealth: facilitators, barriers, and impact of nurse support among high-risk dialysis patients. Telemed J E Health 2013 Aug;19(8):573-578. [doi: 10.1089/tmj.2012.0201] [Medline: 23742629]
59. Berman SJ, Wada C, Minatodani D, Halliday T, Miyamoto R, Lindo J, et al. Home-based preventative care in high-risk dialysis patients: a pilot study. Telemed J E Health 2011 May;17(4):283-287. [doi: 10.1089/tmj.2010.0169] [Medline: 21480788]
60. Sevick MA, Piraino B, Sereika S, Starrett T, Bender C, Bernardini J, et al. A preliminary study of PDA-based dietary self-monitoring in hemodialysis patients. J Ren Nutr 2005 Jul;15(3):304-311. [doi: 10.1016/j.jrn.2005.04.003] [Medline: 16007560]
61. Gallar P, Vigil A, Rodriguez I, Ortega O, Gutierrez M, Hurtado J, et al. Two-year experience with telemedicine in the follow-up of patients in home peritoneal dialysis. J Telemed Telecare 2007;13(6):288-292. [doi: 10.1258/135763307781644906] [Medline: 17785025]
62. Whitten P, Buis L. Use of telemedicine for haemodialysis: perceptions of patients and health-care providers, and clinical effects. J Telemed Telecare 2008;14(2):75-78. [doi: 10.1258/jtt.2007.070411] [Medline: 18348752]
63. Stark S, Snetselaar L, Piraino B, Stone RA, Kim S, Hall B, et al. Personal digital assistant-based self-monitoring adherence rates in 2 dialysis dietary intervention pilot studies: BalanceWise-HD and BalanceWise-PD. J Ren Nutr 2011 Nov;21(6):492-498 [FREE Full text] [doi: 10.1053/j.jrn.2010.10.026] [Medline: 21420316]
64. Connelly K, Siek KA, Chaudry B, Jones J, Astroth K, Welch JL. An offline mobile nutrition monitoring intervention for varying-literacy patients receiving hemodialysis: a pilot study examining usage and usability. J Am Med Inform Assoc 2012;19(5):705-712 [FREE Full text] [doi: 10.1136/amiajnl-2011-000732] [Medline: 22582206]
65. Diamantidis CJ, Fink W, Yang S, Zuckerman MR, Ginsberg J, Hu P, et al. Directed use of the internet for health information by patients with chronic kidney disease: prospective cohort study. J Med Internet Res 2013 Nov 15;15(11):e251 [FREE Full text] [doi: 10.2196/jmir.2848] [Medline: 24240617]
66. Ogna VF, Pruijm M, Zweiacker C, Wuerzner G, Tousset E, Burnier M. Clinical benefits of an adherence monitoring program in the management of secondary hyperparathyroidism with cinacalcet: results of a prospective randomized controlled study. Biomed Res Int 2013;2013:104892 [FREE Full text] [doi: 10.1155/2013/104892] [Medline: 23971019]
67. Heiden S, Buus AA, Jensen MH, Hejlesen OK. A diet management information and communication system to help chronic kidney patients cope with diet restrictions. Stud Health Technol Inform 2013;192:543-547. [doi: 10.3233/978-1-61499-289-9-543] [Medline: 23920614]
68. Neumann CL, Wagner F, Menne J, Brockes C, Schmidt-Weitmann S, Rieken EM, et al. Body weight telemetry is useful to reduce interdialytic weight gain in patients with end-stage renal failure on hemodialysis. Telemed J E Health 2013 Jun;19(6):480-486. [doi: 10.1089/tmj.2012.0188] [Medline: 23614336]
69. Welch JL, Astroth KS, Perkins SM, Johnson CS, Connelly K, Siek KA, et al. Using a mobile application to self-monitor diet and fluid intake among adults receiving hemodialysis. Res Nurs Health 2013 Jun;36(3):284-298 [FREE Full text] [doi: 10.1002/nur.21539] [Medline: 23512869]
70. Blakeman T, Blickem C, Kennedy A, Reeves D, Bower P, Gaffney H, et al. Effect of information and telephone-guided access to community support for people with chronic kidney disease: randomised controlled trial. PLoS One 2014;9(10):e109135 [FREE Full text] [doi: 10.1371/journal.pone.0109135] [Medline: 25330169]
71. Harrington DM, Myers L, Eisenman K, Bhise V, Nayak KS, Rosner MH. The use of a tablet computer platform to optimize the care of patients receiving peritoneal dialysis: a pilot study. Blood Purif 2014;37(4):311-315. [doi: 10.1159/000365440] [Medline: 25170838]
72. Diamantidis CJ, Ginsberg JS, Yoffe M, Lucas L, Prakash D, Aggarwal S, et al. Remote usability testing and satisfaction with a mobile health medication inquiry system in CKD. Clin J Am Soc Nephrol 2015 Aug 7;10(8):1364-1370 [FREE Full text] [doi: 10.2215/CJN.12591214] [Medline: 26220816]
73. van Lint CL, van der Boog PJ, Wang W, Brinkman WP, Rövekamp TJ, Neerincx MA, et al. Patient experiences with self-monitoring renal function after renal transplantation: results from a single-center prospective pilot study. Patient Prefer Adherence 2015;9:1721-1731 [FREE Full text] [doi: 10.2147/PPA.S92108] [Medline: 26673985]
74. Dey V, Jones A, Spalding EM. Telehealth: acceptability, clinical interventions and quality of life in peritoneal dialysis. Sage Open Med 2016;4:2050312116670188 [FREE Full text] [doi: 10.1177/2050312116670188] [Medline: 27757228]
75. Ong SW, Jassal SV, Miller JA, Porter EC, Cafazzo JA, Seto E, et al. Integrating a smartphone-based self-management system into usual care of advanced CKD. Clin J Am Soc Nephrol 2016 Jun 6;11(6):1054-1062 [FREE Full text] [doi: 10.2215/CJN.10681015] [Medline: 27173169]
76. Hayashi A, Yamaguchi S, Waki K, Fujiu K, Hanafusa N, Nishi T, et al. Testing the feasibility and usability of a novel smartphone-based self-management support system for dialysis patients: a pilot study. JMIR Res Protoc 2017 Apr 20;6(4):e63 [FREE Full text] [doi: 10.2196/resprot.7105] [Medline: 28428168]
77. Liu N, Kim J, Jung Y, Arisy A, Nicdao MA, Mikaheal M, et al. Remote monitoring systems for chronic patients on home hemodialysis: field test of a copresence-enhanced design. JMIR Hum Factors 2017 Aug 29;4(3):e21 [FREE Full text] [doi: 10.2196/humanfactors.7078] [Medline: 28851680]

78. Elbert NJ, van Os-Medendorp H, van Renselaar W, Ekeland AG, Hakkaart-van Roijen L, Raat H, et al. Effectiveness and cost-effectiveness of ehealth interventions in somatic diseases: a systematic review of systematic reviews and meta-analyses. J Med Internet Res 2014 Apr 16;16(4):e110 [FREE Full text] [doi: 10.2196/jmir.2790] [Medline: 24739471]
79. Tsai YC, Hung CC, Hwang SJ, Wang SL, Hsiao SM, Lin MY, et al. Quality of life predicts risks of end-stage renal disease and mortality in patients with chronic kidney disease. Nephrol Dial Transplant 2010 May;25(5):1621-1626. [doi: 10.1093/ndt/gfp671] [Medline: 20037172]
80. Mapes DL, Lopes AA, Satayathum S, McCullough KP, Goodkin DA, Locatelli F, et al. Health-related quality of life as a predictor of mortality and hospitalization: the dialysis outcomes and practice patterns study (DOPPS). Kidney Int 2003 Jul;64(1):339-349 [FREE Full text] [doi: 10.1046/j.1523-1755.2003.00072.x] [Medline: 12787427]
81. Kelders SM, Kok RN, Ossebaard HC, van Gemert-Pijnen JE. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012 Nov 14;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]
82. Michie S, Yardley L, West R, Patrick K, Greaves F. Developing and evaluating digital interventions to promote behavior change in health and health care: recommendations resulting from an international workshop. J Med Internet Res 2017 Jun 29;19(6):e232 [FREE Full text] [doi: 10.2196/jmir.7126] [Medline: 28663162]
83. Wade DT. Research into the black box of rehabilitation: the risks of a type III error. Clin Rehabil 2001 Feb;15(1):1-4. [doi: 10.1191/026921501675961253] [Medline: 11237155]
84. van der Meij E, Anema JR, Leclercq WK, Bongers MY, Consten EC, Koops SE, et al. Personalised perioperative care by e-health after intermediate-grade abdominal surgery: a multicentre, single-blind, randomised, placebo-controlled trial. Lancet 2018 Jul 7;392(10141):51-59. [doi: 10.1016/S0140-6736(18)31113-9] [Medline: 29937195]
85. Chen J, Ou L, Hollis SJ. A systematic review of the impact of routine collection of patient reported outcome measures on patients, providers and health organisations in an oncologic setting. BMC Health Serv Res 2013 Jun 11;13:211 [FREE Full text] [doi: 10.1186/1472-6963-13-211] [Medline: 23758898]
86. Jonkman NH, Westland H, Groenwold RH, Ågren S, Anguita M, Blue L, et al. What are effective program characteristics of self-management interventions in patients with heart failure? An individual patient data meta-analysis. J Card Fail 2016 Nov;22(11):861-871 [FREE Full text] [doi: 10.1016/j.cardfail.2016.06.422] [Medline: 27374838]
87. Bos-Touwen I, Jonkman N, Westland H, Schuurmans M, Rutten F, de Wit N, et al. Tailoring of self-management interventions in patients with heart failure. Curr Heart Fail Rep 2015 Jun;12(3):223-235 [FREE Full text] [doi: 10.1007/s11897-015-0259-3] [Medline: 25929690]
88. Duff OM, Walsh DM, Furlong BA, O'Connor NE, Moran KA, Woods CB. Behavior change techniques in physical activity ehealth interventions for people with cardiovascular disease: systematic review. J Med Internet Res 2017 Aug 2;19(8):e281 [FREE Full text] [doi: 10.2196/jmir.7782] [Medline: 28768610]
89. van Velsen L, Wentzel J, van Gemert-Pijnen JE. Designing ehealth that matters via a multidisciplinary requirements development approach. JMIR Res Protoc 2013 Jun 24;2(1):e21 [FREE Full text] [doi: 10.2196/resprot.2547] [Medline: 23796508]
90. Fleuren MA, Paulussen TG, van Dommelen P, van Buuren S. Towards a measurement instrument for determinants of innovations. Int J Qual Health Care 2014 Oct;26(5):501-510 [FREE Full text] [doi: 10.1093/intqhc/mzu060] [Medline: 24951511]
91. Huijg JM, Gebhardt WA, Dusseldorp E, Verheijden MW, van der Zouwe N, Middelkoop BJ, et al. Measuring determinants of implementation behavior: psychometric properties of a questionnaire based on the theoretical domains framework. Implement Sci 2014 Mar 19;9:33 [FREE Full text] [doi: 10.1186/1748-5908-9-33] [Medline: 24641907]
92. Verberne WR, Das-Gupta Z, Allegretti AS, Bart HA, van Biesen W, García-García G, et al. Development of an international standard set of value-based outcome measures for patients with chronic kidney disease: a report of the international consortium for health outcomes measurement (ICHOM) CKD working group. Am J Kidney Dis 2019 Mar;73(3):372-384 [FREE Full text] [doi: 10.1053/j.ajkd.2018.10.007] [Medline: 30579710]
93. Eysenbach G, Consort-eHealth Group. Consort-eHealth: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
94. Liu L, Stroulia E, Nikolaidis I, Miguel-Cruz A, Rios Rincon A. Smart homes and home health monitoring technologies for older adults: a systematic review. Int J Med Inform 2016 Jul;91:44-59. [doi: 10.1016/j.ijmedinf.2016.04.007] [Medline: 27185508]
95. Capurro D, Ganzinger M, Perez-Lu J, Knaup P. Effectiveness of ehealth interventions and information needs in palliative care: a systematic literature review. J Med Internet Res 2014 Mar 7;16(3):e72 [FREE Full text] [doi: 10.2196/jmir.2812] [Medline: 24610324]
96. Fisher E, Law E, Palermo TM, Eccleston C. Psychological therapies (remotely delivered) for the management of chronic and recurrent pain in children and adolescents. Cochrane Database Syst Rev 2015 Mar 23(3):CD011118 [FREE Full text] [doi: 10.1002/14651858.CD011118.pub2] [Medline: 25803793]

97. Chen YR, Schulz PJ. The effect of information communication technology interventions on reducing social isolation in the elderly: a systematic review. J Med Internet Res 2016 Jan 28;18(1):e18 [FREE Full text] [doi: 10.2196/jmir.4596] [Medline: 26822073]
98. Rost T, Stein J, Löbner M, Kersting A, Luck-Sikorski C, Riedel-Heller SG. User acceptance of computerized cognitive behavioral therapy for depression: systematic review. J Med Internet Res 2017 Sep 13;19(9):e309 [FREE Full text] [doi: 10.2196/jmir.7662] [Medline: 28903893]

Abbreviations

BP: blood pressure
CCAT: Crowe Critical Appraisal Tool
CKD: chronic kidney disease
CRD: Centre for Reviews and Dissemination
eHealth: electronic health
GFR: glomerular filtration rate
NCDs: noncommunicable diseases
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QoL: quality of life
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 02.10.18; peer-reviewed by YT Cheung, T Palermo; comments to author 02.02.19; revised version received 28.06.19; accepted 19.08.19; published 05.11.19.

Please cite as:
Shen H, van der Kleij RMJJ, van der Boog PJM, Chang X, Chavannes NH
Electronic Health Self-Management Interventions for Patients With Chronic Kidney Disease: Systematic Review of Quantitative and Qualitative Evidence
J Med Internet Res 2019;21(11):e12384
URL: https://www.jmir.org/2019/11/e12384
doi:10.2196/12384
PMID:31687937

Review

Health Researchers' Use of Social Media: Scoping Review

Justine Dol[1,2], MSc; Perri R Tutelman[2,3], BHSc (Hons); Christine T Chambers[2,3,4], PhD; Melanie Barwick[5,6,7], PhD; Emily K Drake[1], MA; Jennifer A Parker[2], PhD; Robin Parker[8], MLIS; Eric I Benchimol[9,10], MD, PhD; Ronald B George[11], MD; Holly O Witteman[12,13,14,15], PhD

[1]Dalhousie University, Faculty of Health, Halifax, NS, Canada
[2]IWK Health Centre, Centre for Pediatric Pain Research, Halifax, NS, Canada
[3]Dalhousie University, Department of Psychology and Neuroscience, Halifax, NS, Canada
[4]Dalhousie University, Department of Pediatrics, Halifax, NS, Canada
[5]The Hospital for Sick Children, SickKids Research Institute, Child Health Evaluative Sciences, Toronto, ON, Canada
[6]University of Toronto, Faculty of Medicine, Toronto, ON, Canada
[7]University of Toronto, Dalla Lana School of Public Health, Toronto, ON, Canada
[8]Dalhousie University, WK Kellogg Health Sciences Library, Halifax, NS, Canada
[9]University of Ottawa, Faculty of Medicine, Department of Pediatrics, Ottawa, ON, Canada
[10]University of Ottawa, Faculty of Medicine, School of Epidemiology and Public Health, Ottawa, ON, Canada
[11]University of California San Francisco, Department of Anesthesia and Perioperative Care, San Francisco, CA, United States
[12]Laval University, Faculty of Medicine, Department of Family and Emergency Medicine, Quebec, QC, Canada
[13]Laval University, Faculty of Medicine, Office of Education and Professional Development, Quebec, QC, Canada
[14]CHU de Québec-Université Laval, Quebec, QC, Canada
[15]Ottawa Hospital Research Institute, Ottawa, ON, Canada

Corresponding Author:
Christine T Chambers, PhD
Dalhousie University
Department of Psychology and Neuroscience
5850/5980 University Ave
Halifax, NS
Canada
Phone: 1 902 470 8877
Email: christine.chambers@dal.ca

Abstract

Background: Health researchers are increasingly using social media in a professional capacity, and the applications of social media for health researchers are vast. However, there is currently no published evidence synthesis of the ways in which health researchers use social media professionally, and uncertainty remains as to how best to harness its potential.

Objective: This scoping review aimed to explore how social media is used by health researchers professionally, as reported in the literature.

Methods: The scoping review methodology guided by Arksey and O'Malley and Levac et al was used. Comprehensive searches based on the concepts of health research and social media were conducted in MEDLINE, EMBASE, CINAHL, PsycINFO, ERIC, and Web of Science databases, with no limitations applied. Articles were screened at the title and abstract level and at full text by two reviewers. One reviewer extracted data that were analyzed descriptively to map the available evidence.

Results: A total of 8359 articles were screened at the title and abstract level, of which 719 were also assessed at full text for eligibility. The 414 articles identified for inclusion were published in 278 different journals. Studies originated from 31 different countries, with the most prevalent being the United States (52.7% [218/414]). The health discipline of the first authors varied, with medicine (33.3% [138/414]) being the most common. A third of the articles covered health generally, with 61 health-specific topics. Papers used a range of social media platforms (mean 1.33 [SD 0.7]). A quarter of the articles screened reported on social media use for participant recruitment (25.1% [104/414]), followed by practical ways to use social media (15.5% [64/414]), and use of social media for content analysis research (13.3% [55/414]). Articles were categorized as *celebratory* (ie, opportunities

for engagement, 72.2% [299/414]), *contingent* (ie, opportunities and possible limitations, 22.7% [94/414]) and *concerned* (ie, potentially harmful, 5.1% [21/414]).

Conclusions: Health researchers are increasingly publishing on their use of social media for a range of professional purposes. Although most of the sentiment around the use of social media in health research was celebratory, the uses of social media varied widely. Future research is needed to support health researchers to optimize their social media use.

(J Med Internet Res 2019;21(11):e13687) doi:10.2196/13687

KEYWORDS

health; social media; review

Introduction

Health researchers are using social media in a professional capacity [1]. Defined as interactive internet-based applications that enable users to share information, network, and collaborate on the Web [2], well-known examples of social media platforms include Twitter, Facebook, and YouTube. Although social media use has historically been met with skepticism in the health research community [3,4], several major science journals now publish articles endorsing the relevance of social media for researchers [5,6]. The applications of social media for health research are vast; researchers report using social media to post content and keep abreast of advancements in their field (eg, new publications), to network with colleagues and knowledge users (eg, hashtag communities and journal clubs), to conduct research (eg, participant recruitment and social media as a dissemination or data collection tool), and for academic promotion [5,7,8].

Researchers are now actively encouraged to utilize social media in their research, with social media engagement being increasingly recognized by institutions as an important evaluation criterion for promotion and tenure [9]. Social media in the form of live-tweeting is increasingly present at academic conferences, where delegates share content using a specified conference hashtag [10], and many journals now have dedicated editors or committees who promote newly published scientific papers via social media [11]. Evidence is rapidly accumulating on the scholarly impact of social media activities. Some studies have shown that the promotion of research articles over social media channels significantly increases their reach, as evidenced by more article views, downloads, and citations [12-15], though evidence from a randomized controlled trial at a single journal on 243 articles found no difference in page views [16]. Social media is also influencing how we disseminate our research. Many journals have adopted the use of *visual abstracts*, a simple visual representation of the key findings, designed to enhance social media dissemination of health research [17] or encouraging the submission of a tweet to be sent out when the article is published.

Previous reviews have summarized the literature on researchers' use of social media platforms [7] in specific areas of health [18] and for certain purposes [19]. A scoping review was conducted in 2013 on social media use by health professionals and trainees, rather than health researchers [20]. However, there is no evidence synthesis of the ways in which health researchers, as a specific population group, are using social media across platforms, and there remains uncertainty about how to best harness the potential of this medium in health research. A recent review on the use of Twitter to drive research impact concluded that "advice and guidance on the use of social media for research studies is not well understood or exploited by the research community" [21]. Therefore, the objective for this scoping review was to map the literature on the ways in which health researchers report on their use of social media from the existing literature.

Methods

Overview

Health researchers' use of social media was explored using a scoping review guided by the methodology of Arksey and O'Malley [22] and Levac et al [23]. A scoping review protocol was created to guide the process and is available from the corresponding author upon request. This paper adheres to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) extension for scoping reviews [24].

Search Strategy

An experienced information specialist (RP) developed comprehensive search strategies for 6 electronic citation databases: MEDLINE (OvidSP, 1946 to May 2018), EMBASE (Elsevier, 1947 to May 2018), CINAHL (EBSCOHost, 1971 to May 2018), PsycINFO (EBSCOHost, 1967 to May 2018), ERIC (ProQuest, 1966 to May 2018), and Web of Science Core Collection (Clarivate Analytics, 1900 to May 2018). The search strategies utilized index terms, where appropriate, and free text terms to capture the following concepts: (1) social media, including both general terms and specific platform names and terms (eg, Twitter, tweet, Facebook, Snapchat, YouTube); (2) research or researchers; and (3) health or medicine descriptors were added in the non–health discipline databases only (ie, ERIC and Web of Science). The search approach balanced comprehensiveness with precision by including and exploring the general index terms such as *research*, *scientist*, and *social media*, while using adjacency operators to combine the free text search terms. Applying adjacency (or proximity) operators to the text word terms restricts results to those where a relational association exists in the title or abstract text between social media and the research context or researchers. Before finalizing, the searches were checked for sensitivity and relevance and peer reviewed for accuracy and consistency. For the full search strategies in all databases, see Multimedia Appendix 1.

Inclusion and Exclusion Criteria

Articles were determined eligible for inclusion if they discussed the use of social media by health researchers, including but not

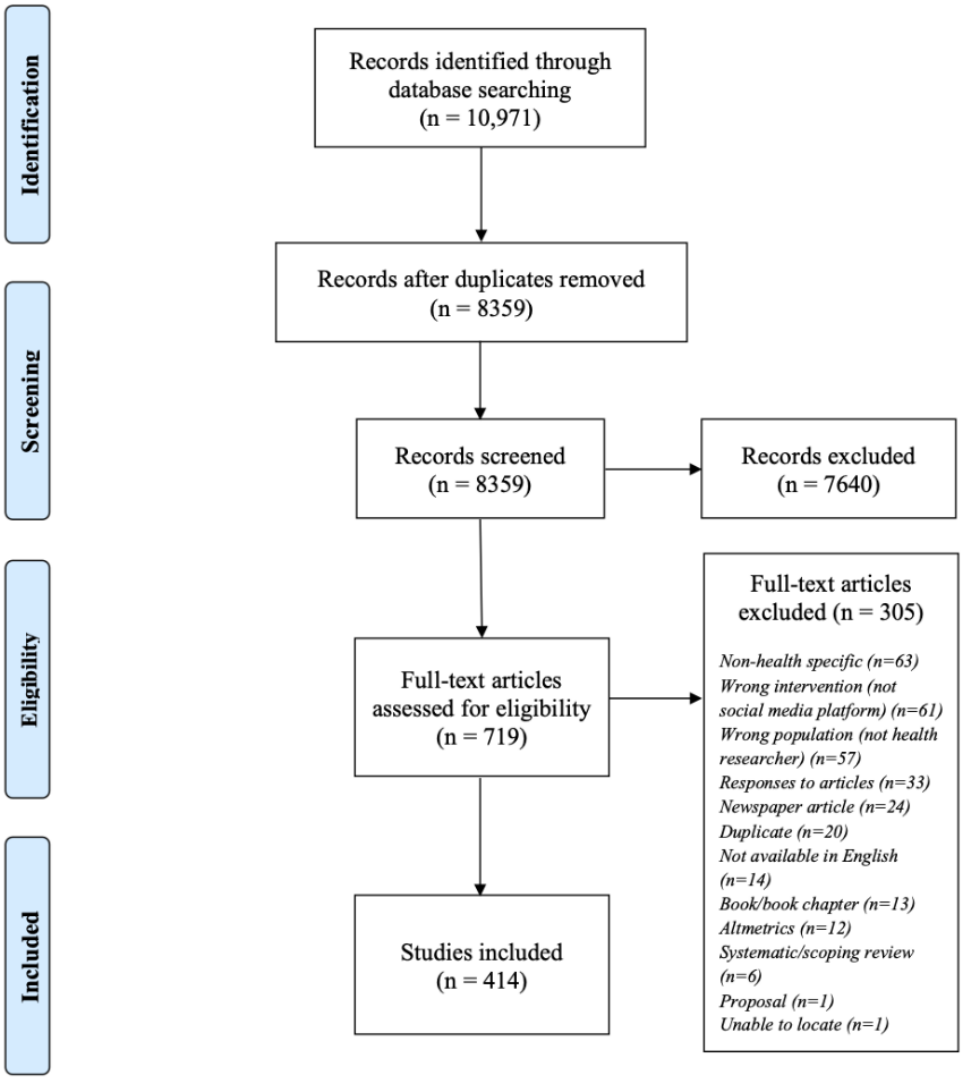
limited to use of social media for recruitment, data mining, social media initiatives or campaigns, hashtag communities, and journal clubs. Articles could be from health researchers at any stage of their career (trainee to faculty member) and across any types of health research (policy, services, outcomes, medical, and basic). Articles were included if authors studied or commented on the uses, benefits, or limitations of social media for health researchers. All article types were included, including dissertations, conference abstracts, and opinion pieces, with the exception of systematic or scoping reviews, books, or book chapters. All articles published since 2000 were included given the rapid advancement of social media and the limited social media literature available before 2000.

Articles were excluded if they were written in a language other than English or if they focused on health care providers' social media use outside of research, organizational, private sector (eg, publishers), or funding agency context. Systematic reviews, scoping reviews, books, and book chapters were excluded. Articles were excluded if they only used social media as a method of recruitment without reporting the uses, benefits, or limitations in relation to their study. Articles were also excluded if they solely reported on the secondary analysis of research output of health researchers, such that a study analyzing research impact data would be excluded (eg, Altmetric reports of a published article).

Data Extraction

The screening process was conducted using the PRISMA extension for scoping reviews (Figure 1) [24,25]. At least two reviewers (JD, PRT, JAP, and EKD) screened the titles and abstracts using Covidence [26]. One reviewer (JD) extracted all data from the included articles using abstract data when available with a standardized Google form that was approved by the project team. Extracted data included article characteristics (year of publication, journal, country of first author, article type, health discipline, and academic affiliation), area of health research, social media platform, and preidentified categories related to the purpose for social media use (eg, recruitment and content analysis).

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses for Scoping Reviews flow diagram of the search and study selection process.



As the internet is now a major source of health information, user sentiment regarding this growth within the medical, sociological, and popular literature has been categorized by Nettleton et al [27] as *celebratory* (offering opportunities of empowerment and engagement for individuals), *contingent* (recognizes the potential positive empowerment yet acknowledges potential limitations), or *concerned* (identified as potentially dangerous owing to unknown quality or reliability of Web-based information). These categories of user sentiment were applied to the identified articles in this review by the reviewer who extracted the data.

For all variables but one, a single response option was selected that best characterized the article. The exception was made for social media platform used, whereby all platforms mentioned in an article were selected. Articles describing general social media use were tagged as such. Extracted data were exported from Google Forms into Microsoft Excel to be cleaned before being imported into IBM SPSS version 22.0 for analysis using descriptive statistics (eg, totals and percentages).

Results

Study Selection

On the basis of the initial search, 8359 articles were identified after duplicates were removed. At the title and abstract screening stage, 7640 articles were excluded. A total of 719 articles were screened as full text, and a further 305 articles were excluded for reasons outlined in Figure 1. One reviewer (JD) extracted all data from the 414 included articles.

Article Characteristics

The study identified 414 unique articles across 278 different journals. The number of articles published on health researchers' use of social media has increased significantly over time (see Figure 2), ranging from 1 publication in 2007 to 88 in 2017. Articles are most commonly published in the *Journal of Medical Internet Research* (6.8% [28/414]) and the JMIR sister journals: *JMIR Research Protocols*, *JMIR Mental Health*, *JMIR Public Health and Surveillance*, and *JMIR Medical Education* (combined 2.4% [10/414]). The next most common journals were *PLoS One* (2.4% [10/414]), *American Journal of Bioethics* (2.2% [9/414]), *AIDS and Behavior* (1.4% [6/414]), and *Nurse Researcher* (1.4% [6/414]). The remaining journals published 5 or fewer social media papers each. Nearly half of the studies published were empirical (42.8% [177/414]), followed by commentaries or opinion pieces (26.3% [109/414]) and conference abstracts (12.6% [52/414]). Other types of papers included discussion papers; theoretical, ethical, or methodological papers; literature reviews; and dissertations.

First authors of included articles represented 31 different countries, most commonly the United States (52.7% [218/414]), the United Kingdom (11.4% [47/414]), Australia (9.4% [39/414]), and Canada (7.0% [29/414]). The remaining countries of origin are shown in Table 1. The health discipline of the first authors varied, with the most common being medicine (33.3% [138/414]), nursing (10.9% [45/414]), public health (7.7% [32/414]), and psychology (3.6% [15/414]). First author discipline was unclear or not specified for 60 articles (14.5% [60/414]), and 30 articles (7.2% [30/414]) pertained to disciplines outside the health field, including but not limited to communication studies, journalism, law, and information studies.

Figure 2. Number of publications by year.

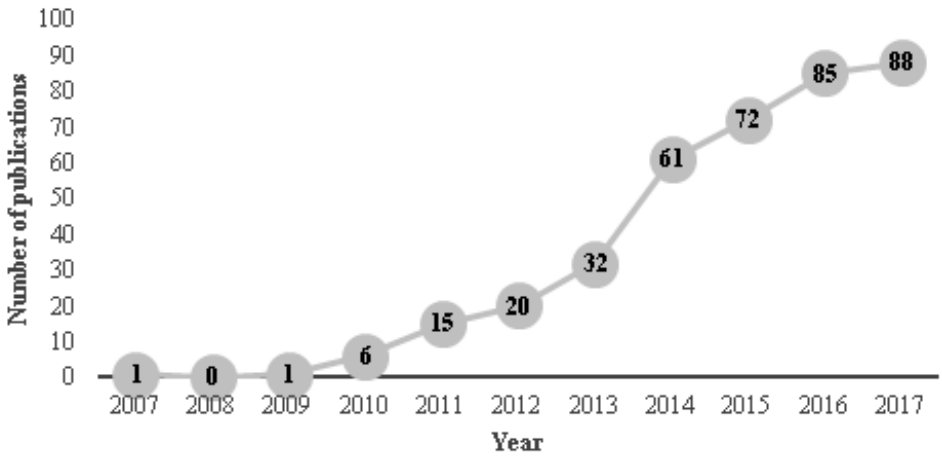


Table 1. Country of first author (N=414).

Country of first author	Value, n (%)
United States	218 (52.7)
United Kingdom	47 (11.4)
Australia	39 (9.4)
Canada	29 (7)
Unknown	28 (6.8)
Germany, Saudi Arabia	6 (1.4)
Brazil, Italy, New Zealand, Spain	4 (1)
Ireland	3 (0.7)
Hong Kong, India, Israel, Norway, the Netherlands	2 (0.5)
Chile, Denmark, Finland, France, Iran, Japan, Northern Ireland, Singapore, South Africa, Sweden, Uganda	1 (0.2)
Multiple countries (cowritten article)	1 (0.2)

Area of Health Research

A third of the articles included were nonspecific and covered health broadly (33.1% [137/414]), touching on 61 different health areas. Predominant topics included infectious diseases (eg, West Nile, Ebola, Zika, and HIV, 7.2% [30/414]), substance use (eg, smoking, alcohol, and marijuana, 6.8% [28/414]), cancer (6.5% [27/414]), mental health (eg, depression and anxiety, 5.3% [22/414]), and chronic disease (eg, diabetes and dementia, 4.6% [19/414]). The remainder of the topics were covered in fewer than 3% of studies.

Purpose for Social Media Use

A quarter of the articles used social media for purposes of participant recruitment (25.1% [104/414]), followed by discussion on practical ways to use social media (15.5% [64/414]) or for content analysis (eg, the frequency and content of tweets on a certain topic; 13.3% [55/414]). Table 2 outlines the full list of uses for social media covered in the papers.

Table 2. Social media purpose (N=414).

Social media purpose	Value, n (%)
Participant recruitment	104 (25.1)
Practical use of social media	64 (15.5)
Content analysis	55 (13.3)
Promotion of academic research	43 (10.4)
Ethics and ethical concerns	33 (8.0)
Data mining from social media	26 (6.3)
Intervention or campaign implementation	26 (6.3)
Engagement of knowledge users	15 (3.6)
Conference tweeting	14 (3.4)
Research education (virtual journal clubs)	12 (2.9)
Data collection from participants	6 (1.4)
Reporting of research findings	4 (1.0)
Accessing scientific resources	4 (1.0)
Crowdfunding	3 (0.7)
Patient education and care	3 (0.7)
Collaborator engagement	1 (0.2)
Information health management	1 (0.2)

Social Media Platforms

In articles that used or discussed at least one specific social media platform, an average of 1.33 (SD 0.7) different platforms were specified; 101 (24.4% [101/414]) articles did not specify a specific social media platform. Of those that did specify a platform, the most common were Twitter (38.2% [158/414]), Facebook (34.8% [144/414]), blogs (8.2% [34/414]), YouTube

(6.3% [26/414/]), LinkedIn (2.7% [11/414]), Instagram (1.9% [8/414]), or research websites, such as ResearchGate or Academia.edu (1.5% for both [6/414]). Other platforms identified were used fewer than 5 times, including but not limited to crowdfunding platforms (eg, GoFundMe, MySpace, Google+, and Pinterest). Table 3 outlines how health researchers are currently using each social media platform for research purposes. For example, most participant recruitment occurs on Facebook (81.7% [335/414]), whereas content analysis (54.5% [226/414]) and data mining (76.9% [318/414]) occurs on Twitter.

Table 3. Social media use by social media platform.

Social media use	Twitter, n	Facebook, n	Blogs[a], n	YouTube, n	LinkedIn, n	Instagram, n	Research websites[b], n	Other platforms, n	Platform not specified, n
Participant recruitment	20	85	4	1	—[c]	1	—	11	11
Practical use of social media	25	13	6	3	6	2	2	2	33
Content analysis	30	7	6	11	—	3	—	3	—
Promotion of academic research	18	7	14	1	2	—	4	1	13
Ethics and ethical concerns	3	7	1	—	1	—	—	—	24
Data mining from social media	20	—	—	4	—	—	—	1	2
Intervention or campaign implementation	10	15	1	3	1	—	—	2	2
Engagement of knowledge users	3	6	1	—	—	1	—	1	5
Conference tweeting	14	—	—	—	—	—	—	—	—
Research education (online journal clubs)	7	—	—	1	—	—	—	—	4
Data collection from participants	2	2	—	—	—	1	—	—	1
Reporting of research findings	3	1	1	—	—	—	—	—	1
Accessing scientific resources	2	—	—	—	—	—	—	—	1
Crowdfunding	—	—	—	—	—	—	—	3	—
Patient education and care	—	—	—	1	—	—	—	—	3
Collaborator engagement	1	1	—	1	—	—	—	—	—
Information health management	—	—	—	—	—	—	—	—	1

[a]WordPress and Tumblr.

[b]eg, ResearchGate.

[c]No data available.

Sentiment Classification on Social Media Use

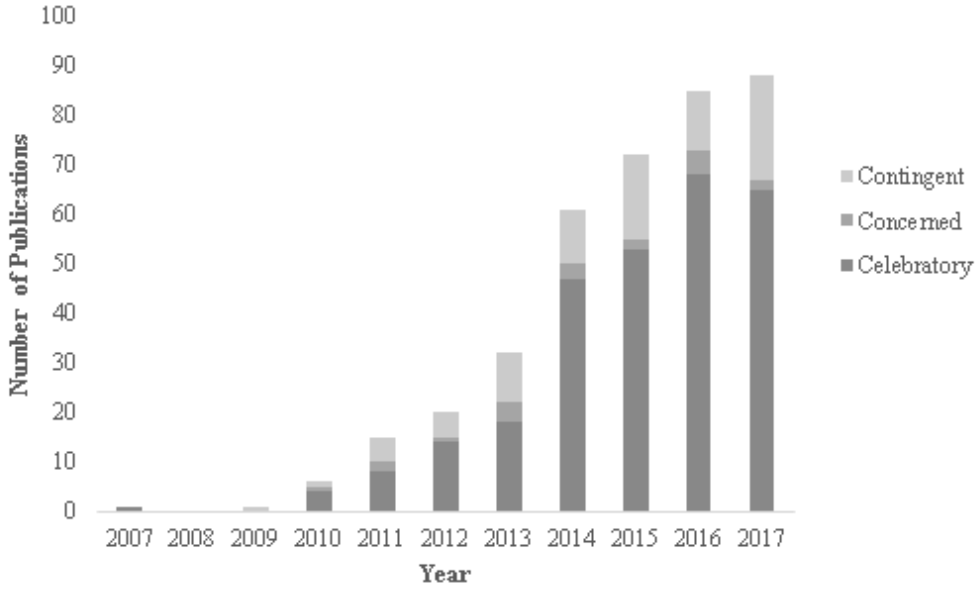
As outlined in Table 4, included articles were most commonly categorized using the Nettleton et al [27] sentiment classification as celebratory, (72.2% [299/414]), followed by contingent (22.7% [94/414]) and concerned (5.1% [21/414]). Articles classified as contingent or concerned were predominantly focused on the ethics of social media or the use of social media for content analysis. With the exception of articles focused on social media ethics (15.2% [63/414]) and use for content analysis (58.2% [241/414]), the remaining articles were predominantly classified as celebratory. Figure 3 plots the classification of articles over time, illustrating the relative consistency of perception over time.

Table 4. Reaction classification of the top 10 social media topics covered.

Social media topics	Celebratory, n (%)	Concerned, n (%)	Contingent, n (%)
Participant recruitment (n=104)	81 (77.9)	4 (3.8)	19 (18.3)
Practical use of social media (n=60)	48 (80)	2 (3)	10 (17)
Content analysis (n=55)	32 (58)	7 (13)	16 (29)
Promotion of academic research (n=43)	32 (77)	2 (5)	9 (21)
Ethics and ethical concerns (n=33)	5 (15)	5 (15)	23 (70)
Data mining from social media (n=26)	22 (85)	—[a]	4 (15)
Intervention or campaign implementation (n=26)	23 (89)	—	3 (11)
Engagement of knowledge users (n=17)	14 (82)	1 (6)	2 (12)
Conference tweeting (n=14)	12 (86)	—	2 (14)
Research education (online journal clubs; n=12)	11 (92)	—	1 (8)

[a]No data available.

Figure 3. Classification of publications by year.



Discussion

Principal Findings and Comparison With Prior Work

Results of this scoping review identified how health researchers are using social media for research purposes within 414 articles that met inclusionary criteria and were published after 2000 and before May 2018, with the first articles on this topic published in 2007. There has been substantial growth in the number of studies published on health researchers' use of social media over the past decade, with the greatest increase occurring over the past 5 years. The increased interest in this field may reflect the widespread adoption of social media across academic contexts, including interest by journals [11], conferences [28], within institutions [9], and among individual scientists [29]. The number of published papers on social media will likely continue to rise, given its increasing popularity within academia.

The vast majority of publications originate from scholars in high-income countries, with 80.4% (333/414) of first authors based in the United States, the United Kingdom, Australia, or Canada. This is consistent with statistics on worldwide social media penetration that show North America ranking first with a social media penetration rate of 70%, followed by Northern Europe with 66% penetration [30]. The global average penetration rate reported in 2018 was 42%. A recent review found that the use of social media for health-related purposes is increasing in low- and middle-income countries [31], yet reliable access, cost, and infrastructure remain a barrier to the internet in some low- and middle-income countries. With regard to possible publication bias, an early study by Man et al [32] identified that research spending and English proficiency were associated with a greater likelihood of having publication output in high-ranking medical journals, which may have influenced the low number of articles identified from low- and middle-income countries. The higher rate of social media publications stemming from North America was also observed in a review published in 2014, in which the authors expected to see a rise in studies from low- and middle-income countries *in the future* [18]. However, it appears that the landscape has not yet shifted. This is intriguing, given that one of the main

advantages of social media lies in its potential to disseminate research evidence with greater reach and at a relatively low cost [33]. Future research should explore the use, opportunities, and barriers of social media use by health researchers based on available technology (ie, bandwidth and hardware), research funding structures, geography, and socioeconomic factors (eg, gender, race, education, and income).

This study revealed that health researchers use social media for a range of research-related purposes. Most commonly, social media is used to recruit participants and to source data from the Web (eg, content analysis of social media posts and data mining on social media). Social media appears to facilitate research on clinical populations who have traditionally been difficult to recruit or study because of stigmatization, social disadvantage, low disease prevalence, or mobility challenges that make physical participation difficult [34]. In this review, we see this reflected in the health areas most associated with social media use, including infectious disease, substance use, cancer, mental health, and chronic disease. The other topics covered in the papers identified were practical use of social media and ethical concerns related to social media use. Articles identified various practical ways to use social media, including but not limited to how to incorporate social media into clinical trials, how to use social media to advance careers, how to use social media to disseminate research findings, and how to use various social media platforms. Similarly, ethical concerns arose related to various topics, including but not limited to social media in clinical trials, privacy concerns, professional relationships, and use of social media as a recruitment tool.

Most of the studies identified in this review used social media in primarily passive ways (eg, for participant recruitment or content analysis), with more active application noted in a handful of studies where it was used to support intervention delivery, promote campaigns (6.3% [26/414]), or build knowledge user engagement (3.6% [15/414]). Health researchers are not yet harnessing the full range of benefits available through social media, and this may reflect their lack of being social media savvy with regard to platform functions, audiences, features, or best practices. Health researchers are also using social media channels to promote their research, a practice that has been associated with increased article views and downloads [35]. Whether promoting articles over social media translates to an increase in citations remains unclear [14,36,37], as does the relationship between social media promotion and traditional academic metrics [38].

Health researchers rarely specify the social media platform used for research purposes, but when they do, they favor Twitter and Facebook. This is largely consistent with what has been reported previously [39-41]; however, this could be related to a function of historical emergence, whereby platforms such as Instagram and Snapchat became available later, and their use has not yet been widely reported in peer-reviewed literature. Health research using these newer platforms is still evolving, and it remains to be seen if it will prove to be useful for research purposes; fewer than 2% of the studies in this review discussed Instagram and none mentioned Snapchat. There is no *one-size-fits-all* platform for research-based social media use because use depends on purpose [42]. For example, Twitter may be effective for connecting with other scientists, but researchers seeking to reach adolescents and young adults should consider YouTube, Instagram, and Snapchat, as these are currently the most popular platforms among this age group [42]. The rapidly evolving social media landscape poses a challenge for health researchers given the significantly slower pace of empirical research and evaluation; some platforms that were popular just a few years ago (eg, MySpace) are now obsolete. This challenge is common among electronic health (eHealth) tools more generally [43] and not unique to social media. Baker et al [44] provide recommendations for conducting eHealth research, including social media, to optimize its timeliness and, in turn, its usefulness and effectiveness.

Health researcher sentiment regarding social media use was mostly celebratory, with a smaller percentage of researchers reporting concern and hesitation related to the ethical use of social media for research and analysis of social media content from online forums, Facebook groups, and Twitter hashtag conversations and comments. These apprehensions are not new; the ethics of conducting research using online communities has been a contentious area of debate over the past two decades [45]. The controversy lies in whether social media content is public or private information and extends to issues around confidentiality, informed consent, voluntary participation, and the potential for harm for both vulnerable populations and for researchers [3,46-48]. For health researchers who are also regulated professionals, social media may pose additional challenges related to patient privacy, maintaining professional boundaries, and the potential for misinterpretation of medical information [49]. In response, some organizations have developed policies to guide social media use among researchers [50-52]. However, there is no common standard policy or procedure to guide social media use in health research.

Limitations

Although this scoping review was conducted according to scoping review methodology, there were some limitations that are worth noting. Data were extracted by only 1 reviewer owing to the high number of studies identified in the data extraction phase. To minimize error, 2 reviewers identified relevant studies, and a standardized extraction form was used to ensure accuracy. As the data extracted were descriptive and did not include study results, the impact of potential data extraction errors is minimal.

We defined health researcher broadly, including the spectrum of bench to bedside. This resulted in a wide search strategy yielding over 8000 initial titles to screen. Social media practices likely vary, such that clinical health researchers working with vulnerable populations on the Web may be more inclined to recognize ethical challenges as compared with a public health researcher seeking to disseminate evidence on the latest flu vaccination. However, broad inclusion of a range of health researchers enabled us to gain a wide-ranging picture of current social media practices, thereby increasing the external validity of our findings.

A final limitation is the rapid growth of the field, whereby challenges with currency of publication is noted. This field is moving rapidly, so it is important to acknowledge that this scoping review is a snapshot at a particular point in time.

Conclusions

In conclusion, health researchers are increasingly using social media for a range of professional purposes, and the evidence reflecting this use varies widely. Although most of the sentiment around the use of social media in health research is celebratory, there are concerns about the ethics of social media use for some purposes. Future initiatives are needed to support health researchers to navigate the social media landscape and evaluate the impact of their efforts. Given the concerns related to ethics and content analysis of social media, future work should focus on providing additional direction to health researchers on how to ethically use and engage with social media. This could include the development of professional or institution-specific guidelines or the development of best practices.

Acknowledgments

Funding support for this research was provided through the Nova Scotia Health Research Foundation – Knowledge Sharing Support Award ($10,000; PSO-KSSA-2016-964). CTC is the senior author for this publication. CTC holds a Tier 1 Canada Research Chair and infrastructure support was provided by the Canada Foundation for Innovation. CTC's research is also supported by the Canadian Institutes of Health Research. PRT is supported by a CIHR Vanier Canada Graduate Scholarship. EIB was supported by a New Investigator Award from the Canadian Institutes of Health Research, Canadian Association of Gastroenterology and Crohn's and Colitis Canada, and by the Career Enhancement Program of the Canadian Child Health Clinician Scientist Program. HOW is supported by a Fonds de recherche du Québec—Santé Research Scholar Junior 2 Award. Article processing fees for this manuscript were provided by CTC, RBG, and HOW. The funders had no involvement in review or approval of the manuscript.

Authors' Contributions

CTC had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. JD, PRT, JAP, and EKD were involved in screening studies. JD and PRT were involved in full text screening, and JD was responsible for data extraction. All authors were involved in the design and conduct of the study and preparation and review of the manuscript.

Conflicts of Interest

ED runs her own consulting business. There are no other conflicts of interest to declare.

Multimedia Appendix 1
Search strategies for electronic databases.
[PDF File (Adobe PDF File), 105 KB - jmir_v21i11e13687_app1.pdf]

References

1. Collins K, Shiffman D, Rock J. How are scientists using social media in the workplace? PLoS One 2016;11(10):e0162680 [FREE Full text] [doi: 10.1371/journal.pone.0162680] [Medline: 27732598]
2. Kaplan AM, Haenlein M. Users of the world, unite! The challenges and opportunities of social media. Bus Horiz 2010;53(1):59-68. [doi: 10.1016/j.bushor.2009.09.003]
3. Bender JL, Cyr AB, Arbuckle L, Ferris LE. Ethics and privacy implications of using the internet and social media to recruit participants for health research: a privacy-by-design framework for online recruitment. J Med Internet Res 2017 Apr 6;19(4):e104 [FREE Full text] [doi: 10.2196/jmir.7029] [Medline: 28385682]
4. Vogel L. Medicine slow to recognize social media as window into the patient experience. Can Med Assoc J 2019 Jan 21;191(3):E87-E88 [FREE Full text] [doi: 10.1503/cmaj.109-5701] [Medline: 30665982]
5. Van Noorden R. Online collaboration: scientists and the social network. Nature 2014 Aug 14;512(7513):126-129. [doi: 10.1038/512126a] [Medline: 25119221]
6. Savage N. Scientists in the Twitterverse. Cell 2015 Jul 16;162(2):233-234 [FREE Full text] [doi: 10.1016/j.cell.2015.06.062] [Medline: 26186181]
7. Sinnenberg L, Buttenheim AM, Padrez K, Mancheno C, Ungar L, Merchant RM. Twitter as a tool for health research: a systematic review. Am J Public Health 2017 Jan;107(1):e1-e8. [doi: 10.2105/AJPH.2016.303512] [Medline: 27854532]
8. Coyne J. Coyne of the Realm. 2013 Dec 9. Advice to Junior Academics on How to Get Involved With Twitter URL: http://www.coyneoftherealm.com/2013/12/09/advice-to-junior-academics-on-how-to-get-involved-with-twitter/ [accessed 2019-09-24]
9. Cabrera D. Mayo Clinic Social Media Network. 2016 May 25. Mayo Clinic includes Social Media Scholarship Activities in Academic Advancement URL: https://socialmedia.mayoclinic.org/2016/05/25/mayo-clinic-includes-social-media-scholarship-activities-in-academic-advancement/ [accessed 2019-10-01]
10. Djuricich AM, Zee-Cheng JE. Live tweeting in medicine: 'Tweeting the meeting'. Int Rev Psychiatry 2015 Apr;27(2):133-139. [doi: 10.3109/09540261.2014.1000270] [Medline: 25739015]

11. Kelly BS, Redmond CE, Nason GJ, Healy GM, Horgan NA, Heffernan EJ. The use of twitter by radiology journals: an analysis of twitter activity and impact factor. J Am Coll Radiol 2016 Nov;13(11):1391-1396. [doi: 10.1016/j.jacr.2016.06.041] [Medline: 27577594]
12. Allen HG, Stanton TR, Di Pietro F, Moseley GL. Social media release increases dissemination of original articles in the clinical pain sciences. PLoS One 2013;8(7):e68914 [FREE Full text] [doi: 10.1371/journal.pone.0068914] [Medline: 23874810]
13. Peoples BK, Midway SR, Sackett D, Lynch A, Cooney PB. Twitter predicts citation rates of ecological research. PLoS One 2016;11(11):e0166570 [FREE Full text] [doi: 10.1371/journal.pone.0166570] [Medline: 27835703]
14. Eysenbach G. Can tweets predict citations? Metrics of social impact based on Twitter and correlation with traditional metrics of scientific impact. J Med Internet Res 2011 Dec 19;13(4):e123 [FREE Full text] [doi: 10.2196/jmir.2012] [Medline: 22173204]
15. Cosco TD. Medical journals, impact and social media: an ecological study of the Twittersphere. Can Med Assoc J 2015 Dec 8;187(18):1353-1357 [FREE Full text] [doi: 10.1503/cmaj.150976] [Medline: 26644544]
16. Fox CS, Bonaca MA, Ryan JJ, Massaro JM, Barry K, Loscalzo J. A randomized trial of social media from Circulation. Circulation 2015 Jan 6;131(1):28-33. [doi: 10.1161/CIRCULATIONAHA.114.013509] [Medline: 25406308]
17. Ibrahim AM, Lillemoe KD, Klingensmith ME, Dimick JB. Visual abstracts to disseminate research on social media: a prospective, case-control crossover study. Ann Surg 2017 Dec;266(6):e46-e48. [doi: 10.1097/SLA.0000000000002277] [Medline: 28448382]
18. Capurro D, Cole K, Echavarría MI, Joe J, Neogi T, Turner AM. The use of social networking sites for public health practice and research: a systematic review. J Med Internet Res 2014 Mar 14;16(3):e79 [FREE Full text] [doi: 10.2196/jmir.2679] [Medline: 24642014]
19. Antheunis ML, Tates K, Nieboer TE. Patients' and health professionals' use of social media in health care: motives, barriers and expectations. Patient Educ Couns 2013 Sep;92(3):426-431. [doi: 10.1016/j.pec.2013.06.020] [Medline: 23899831]
20. Hamm MP, Chisholm A, Shulhan J, Milne A, Scott SD, Klassen TP, et al. Social media use by health care professionals and trainees: a scoping review. Acad Med 2013 Sep;88(9):1376-1383. [doi: 10.1097/ACM.0b013e31829eb91c] [Medline: 23887004]
21. Schnitzler K, Davies N, Ross F, Harris R. Using Twitter™ to drive research impact: a discussion of strategies, opportunities and challenges. Int J Nurs Stud 2016 Jul;59:15-26. [doi: 10.1016/j.ijnurstu.2016.02.004] [Medline: 27222446]
22. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol Theory Pract 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
23. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. Implement Sci 2010 Sep 20;5:69 [FREE Full text] [doi: 10.1186/1748-5908-5-69] [Medline: 20854677]
24. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 2;169(7):467-473. [doi: 10.7326/M18-0850] [Medline: 30178033]
25. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Br Med J 2009 Jul 21;339:b2535 [FREE Full text] [doi: 10.1136/bmj.b2535] [Medline: 19622551]
26. Covidence Systematic Review Software. URL: http://www.covidence.org [accessed 2019-10-01]
27. Nettleton S, Burrows R, O'Malley L. The mundane realities of the everyday lay use of the internet for health, and their consequences for media convergence. Sociol Health Illn 2005 Nov;27(7):972-992 [FREE Full text] [doi: 10.1111/j.1467-9566.2005.00466.x] [Medline: 16313525]
28. Neill A, Cronin JJ, Brannigan D, O'Sullivan R, Cadogan M. The impact of social media on a major international emergency medicine conference. Emerg Med J 2014 May;31(5):401-404. [doi: 10.1136/emermed-2012-202039] [Medline: 23423992]
29. Chambers C. Children's Healthcare Canada. 2013. It Doesn't Have to Hurt: Strategies for Helping Children with Shots and Needles. A YouTube Video for Parents URL: https://www.blog.childrenshealthcarecanada.ca/blog/2013/11/19/20131119it-doesnt-have-to-hurt-strategies-for-helping-children-with-shots-and-needles-a-youtube-video-for-parents [accessed 2019-10-01]
30. Clement J. Statista. 2019. Global Social Network Penetration Rate as of January 2019, by Region URL: https://www.statista.com/statistics/269615/social-network-penetration-by-region/ [accessed 2019-10-01]
31. Hagg E, Dahinten VS, Currie LM. The emerging use of social media for health-related purposes in low and middle-income countries: a scoping review. Int J Med Inform 2018 Jul;115:92-105. [doi: 10.1016/j.ijmedinf.2018.04.010] [Medline: 29779724]
32. Man JP, Weinkauf JG, Tsang M, Sin JH. Why do some countries publish more than others? An international comparison of research funding, English proficiency and publication output in highly ranked general medical journals. Eur J Epidemiol 2004;19(8):811-817. [doi: 10.1023/B:EJEP.0000036571.00320.b8]
33. Moorhead SA, Hazlett DE, Harrison L, Carroll JK, Irwin A, Hoving C. A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. J Med Internet Res 2013 Apr 23;15(4):e85 [FREE Full text] [doi: 10.2196/jmir.1933] [Medline: 23615206]

34. Whitaker C, Stevelink S, Fear N. The use of Facebook in recruiting participants for health research purposes: a systematic review. J Med Internet Res 2017 Aug 28;19(8):e290 [FREE Full text] [doi: 10.2196/jmir.7071] [Medline: 28851679]
35. Gates A, Featherstone R, Shave K, Scott SD, Hartling L. Dissemination of evidence in paediatric emergency medicine: a quantitative descriptive evaluation of a 16-week social media promotion. BMJ Open 2018 Jun 6;8(6):e022298 [FREE Full text] [doi: 10.1136/bmjopen-2018-022298] [Medline: 29880576]
36. de Winter JC. The relationship between tweets, citations, and article views for PLOS ONE articles. Scientometrics 2015;102(2):1773-1779. [doi: 10.1007/s11192-014-1445-x]
37. Haustein S, Peters I, Sugimoto CR, Thelwall M, Larivière V. Tweeting biomedicine: an analysis of tweets and citations in the biomedical literature. J Assn Inf Sci Technol 2014;65(4):656-669. [doi: 10.1002/asi.23101]
38. Wang X, Cui Y, Li Q, Guo X. Social media attention increases article visits: an investigation on article-level referral data of PeerJ. Front Res Metr Anal 2017;2:11. [doi: 10.3389/frma.2017.00011]
39. O'Connor A, Jackson L, Goldsmith L, Skirton H. Can I get a retweet please? Health research recruitment and the Twittersphere. J Adv Nurs 2014 Mar;70(3):599-609. [doi: 10.1111/jan.12222] [Medline: 23909740]
40. van Eperen L, Marincola FM. How scientists use social media to communicate their research. J Transl Med 2011 Nov 15;9:199 [FREE Full text] [doi: 10.1186/1479-5876-9-199] [Medline: 22085450]
41. Gruzd A, Goertzen M. Wired Academia: Why Social Science Scholars are using Social Media. In: Proceedings of the 2013 46th Hawaii International Conference on System Sciences. 2013 Presented at: HICSS'13; January 7-10, 2013; Wailea, Maui, HI, USA p. 3332-3341 URL: https://ieeexplore.ieee.org/document/6480244 [doi: 10.1109/HICSS.2013.614]
42. Smith A, Anderson M. Pew Research Centre. 2018 Mar 1. Social Media Use in 2018 URL: http://www.pewinternet.org/2018/03/01/social-media-use-in-2018/ [accessed 2019-10-01]
43. Higgins KS, Tutelman PR, Chambers CT, Witteman HO, Barwick M, Corkum P, et al. Availability of researcher-led eHealth tools for pain assessment and management: barriers, facilitators, costs, and design. Pain Rep 2018 Sep;3(Suppl 1):e686 [FREE Full text] [doi: 10.1097/PR9.0000000000000686] [Medline: 30324177]
44. Baker TB, Gustafson DH, Shah D. How can research keep up with eHealth? Ten strategies for increasing the timeliness and usefulness of eHealth research. J Med Internet Res 2014 Feb 19;16(2):e36 [FREE Full text] [doi: 10.2196/jmir.2925] [Medline: 24554442]
45. Eysenbach G, Till JE. Ethical issues in qualitative research on internet communities. Br Med J 2001 Nov 10;323(7321):1103-1105 [FREE Full text] [doi: 10.1136/bmj.323.7321.1103] [Medline: 11701577]
46. Kleinsman J, Buckley S. Facebook study: a little bit unethical but worth it? J Bioeth Inq 2015 Jun;12(2):179-182. [doi: 10.1007/s11673-015-9621-0] [Medline: 25740760]
47. Lunnay B, Borlagdan J, McNaughton D, Ward P. Ethical use of social media to facilitate qualitative research. Qual Health Res 2015 Jan;25(1):99-109. [doi: 10.1177/1049732314549031] [Medline: 25212856]
48. Golder S, Ahmed S, Norman G, Booth A. Attitudes toward the ethics of research using social media: a systematic review. J Med Internet Res 2017 Jun 6;19(6):e195 [FREE Full text] [doi: 10.2196/jmir.7082] [Medline: 28588006]
49. Ventola CL. Social media and health care professionals: benefits, risks, and best practices. Pharm Ther 2014 Jul;39(7):491-520 [FREE Full text] [Medline: 25083128]
50. University of York. Social Media Guidelines for Researchers URL: https://www.york.ac.uk/staff/research/governance/research-policies/social-media/ [accessed 2019-10-01]
51. University of Rochester. 2016. Guideline for Research Using Social Media URL: https://www.rochester.edu/ohsp/documents/ohsp/pdf/policiesAndGuidance/Guideline_for_Research_Using_Social_Media.pdf [accessed 2019-10-01]
52. National Institutes of Health. 2016. Guidance Regarding Social Media Tools URL: https://www.nih.gov/health-information/nih-clinical-research-trials-you/guidance-regarding-social-media-tools [accessed 2019-10-01]

Abbreviations

eHealth: electronic health
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by G Eysenbach; submitted 12.02.19; peer-reviewed by M El Tantawi, E Fernandes de Mello Araujo, A Burke; comments to author 28.03.19; revised version received 16.05.19; accepted 29.06.19; published 13.11.19.

Please cite as:
Dol J, Tutelman PR, Chambers CT, Barwick M, Drake EK, Parker JA, Parker R, Benchimol EI, George RB, Witteman HO
Health Researchers' Use of Social Media: Scoping Review
J Med Internet Res 2019;21(11):e13687
URL: https://www.jmir.org/2019/11/e13687
doi:10.2196/13687
PMID:31719028

Review

Accuracy of Wristband Fitbit Models in Assessing Sleep: Systematic Review and Meta-Analysis

Shahab Haghayegh[1], MS; Sepideh Khoshnevis[1], MD, PhD; Michael H Smolensky[1,2], PhD; Kenneth R Diller[1], SCD; Richard J Castriotta[3], MD, FAASM, FCCP

[1]Department of Biomedical Engineering, Cockrell School of Engineering, The University of Texas at Austin, Austin, TX, United States

[2]Division of Pulmonary and Sleep Medicine, Department of Internal Medicine, McGovern School of Medicine, The University of Texas Health Science Center at Houston, Houston, TX, United States

[3]Division of Pulmonary, Critical Care and Sleep Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA, United States

Corresponding Author:
Shahab Haghayegh, MS
Department of Biomedical Engineering
Cockrell School of Engineering
The University of Texas at Austin
107 W Dean Keeton St
Austin, TX,
United States
Phone: 1 5129543436
Email: shahab@utexas.edu

Abstract

Background: Wearable sleep monitors are of high interest to consumers and researchers because of their ability to provide estimation of sleep patterns in free-living conditions in a cost-efficient way.

Objective: We conducted a systematic review of publications reporting on the performance of wristband *Fitbit* models in assessing sleep parameters and stages.

Methods: In adherence with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement, we comprehensively searched the Cumulative Index to Nursing and Allied Health Literature (CINAHL), Cochrane, Embase, MEDLINE, PubMed, PsycINFO, and Web of Science databases using the keyword *Fitbit* to identify relevant publications meeting predefined inclusion and exclusion criteria.

Results: The search yielded 3085 candidate articles. After eliminating duplicates and in compliance with inclusion and exclusion criteria, 22 articles qualified for systematic review, with 8 providing quantitative data for meta-analysis. In reference to polysomnography (PSG), nonsleep-staging *Fitbit* models tended to overestimate total sleep time (TST; range from approximately 7 to 67 mins; effect size=-0.51, $P<.001$; heterogenicity: $I^2=8.8\%$, $P=.36$) and sleep efficiency (SE; range from approximately 2% to 15%; effect size=-0.74, $P<.001$; heterogenicity: $I^2=24.0\%$, $P=.25$), and underestimate wake after sleep onset (WASO; range from approximately 6 to 44 mins; effect size=0.60, $P<.001$; heterogenicity: $I^2=0\%$, $P=.92$) and there was no significant difference in sleep onset latency (SOL; $P=.37$; heterogenicity: $I^2=0\%$, $P=.92$). In reference to PSG, nonsleep-staging *Fitbit* models correctly identified sleep epochs with accuracy values between 0.81 and 0.91, sensitivity values between 0.87 and 0.99, and specificity values between 0.10 and 0.52. Recent-generation *Fitbit* models that collectively utilize heart rate variability and body movement to assess sleep stages performed better than early-generation nonsleep-staging ones that utilize only body movement. Sleep-staging *Fitbit* models, in comparison to PSG, showed no significant difference in measured values of WASO ($P=.25$; heterogenicity: $I^2=0\%$, $P=.92$), TST ($P=.29$; heterogenicity: $I^2=0\%$, $P=.98$), and SE ($P=.19$) but they underestimated SOL ($P=.03$; heterogenicity: $I^2=0\%$, $P=.66$). Sleep-staging *Fitbit* models showed higher sensitivity (0.95-0.96) and specificity (0.58-0.69) values in detecting sleep epochs than nonsleep-staging models and those reported in the literature for regular wrist actigraphy.

Conclusions: Sleep-staging *Fitbit* models showed promising performance, especially in differentiating wake from sleep. However, although these models are a convenient and economical means for consumers to obtain gross estimates of sleep parameters and time spent in sleep stages, they are of limited specificity and are not a substitute for PSG.

(J Med Internet Res 2019;21(11):e16273) doi:10.2196/16273

KEYWORDS

Fitbit; polysomnography; sleep tracker; wearable; actigraphy; sleep diary; sleep stages; accuracy; validation; comparison of performance

Introduction

Polysomnography (PSG) consists of simultaneous electroencephalographic (EEG), electromyographic, electrooculographic, electrocardiographic, and other assessments. PSG is regarded as the gold standard for diagnosis of sleep disorders and conduct of sleep research. However, the environment and instrumentation of conventional PSG can be uncomfortable, anxiety producing, and even sleep disturbing. Additionally, PSG requires a special facility plus oversight by skilled technicians, making it expensive and precluding, under most circumstances, investigation of between-night variation of sleep quality. Thus, it is not surprising less than 50% of sleep studies nowadays are conducted in formal sleep facilities [1].

Sleep diary methods are simple and economical ways of tracking and appraising sleep by consumers but because they entail subjective self-ratings, they are often inaccurate and incomplete; furthermore, they do not assess sleep architecture and stages. EEG wearables enable at-home evaluation of sleep architecture and staging but they are expensive and somewhat technologically complicated. Wrist actigraphy, which senses accelerated motion, was introduced some 35 years ago by Ambulatory Monitoring Inc and is now used in conjunction with proprietary interpretative algorithms to conduct outpatient sleep screenings through estimation of key sleep parameters. Nonetheless, these devices, which rely entirely on movement-based algorithms [2], lack sleep-stage assessment capability; additionally, they tend to overestimate sleep duration [3]. Approximately 10 years ago, *Fitbit* (Fitbit, Inc) introduced its first wearable model [4] for use by health-conscious consumers. Early-generation Fitbit models only determined sleep parameters. However, subsequent modifications and refinements, including a scoring algorithm based collectively on body movement and heart rate variability (HRV), enable recent-generation Fitbit models—Fitbit Charge 2, Fitbit Charge 3, Fitbit Alta HR, Fitbit Versa, Fitbit Versa 2, Fitbit Blaze, Fitbit Inspire HR, and Fitbit Ionic—to estimate not only sleep parameters and stages [5], but wake- and sleep-time heart rate [6]. A 2019 survey found wearable technology to be the number one fitness trend worldwide [7]. Fitbit wearables, in particular, are very popular among consumers, with more than 25 million active users in more than 80 countries [8]. Additionally, they are the most-used wearables for conducting biomedical research [9]. In this regard, this year the United States National Institutes of Health announced its decision to incorporate Fitbit technology into its *All of Us Research Program* [10,11]. Nonetheless, the accuracy of Fitbit technology remains a major concern, not only of medical professionals but the lay public [12]. In recognition of the growing interest and use of personal wristband devices to routinely self-assess biomarkers of sleep quality, the National Sleep Foundation, the Consumer Technology Association, and the American National Standards Institute developed recommended terminology and definitions to describe sleep features derived by such products [13]. Given the growing popularity with consumers and medical organizations of the Fitbit wristband devices, the objective of this paper is to appraise, in compliance with these recommendations, the performance of both early- and recent-generation Fitbit models in determining sleep parameters and sleep stages through a systematic review of findings of relevant publications and a meta-analysis of reported data.

Methods

This prospective systematic review, which was not registered beforehand, was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement [14].

Search Strategy

An online comprehensive search of the following databases was performed: the Cumulative Index to Nursing and Allied Health Literature (CINAHL), the Cochrane Database, Embase, MEDLINE, PsycINFO, PubMed, and Web of Science. The search was initially performed during July 2018 and again during July and October 2019 using the keyword *Fitbit* without language, publication date, or other filters.

Eligibility Criteria

Retrieved publications qualified for the systematic review if they (1) involved validity of sleep data of any marketed Fitbit model and (2) incorporated PSG, actigraphy, home EEG, sleep diary, or survey method as reference. Exclusion criteria included the following: (1) sample size of less than 5 participants, (2) review paper, (3) absent or inappropriate statistical analysis, and (4) duplicate publication of the same data and findings.

Study Selection

Citations were imported into the reference manager software Mendeley. After elimination of duplicate reports, one author (SH) screened titles and abstracts first to remove unrelated publications; thereafter, two authors (SH and SK) independently screened remaining publications for eligibility according to inclusion and exclusion criteria. Disagreements were resolved by discussion.

Data Extraction and Items

The following items were extracted in a systematic manner by one author (SH) and checked for accuracy by another author (SK): first author; year of publication; type of sleep tracker and comparator; number, sex, type, and age of participants; study site; number of nights of sleep assessment; bedtime; Fitbit mode setting; anatomical placement of tracker; and study outcomes relative to the denoted reference standard—the precision of measuring the parameters of total sleep time (TST), sleep onset latency (SOL), wake after sleep onset (WASO), and sleep

efficiency (SE), as well as the sensitivity, specificity, and accuracy of detecting both sleep epochs and sleep stages.

Bias Assessment

A checklist, adapted from Downs and Black [15], was applied to evaluate each publication for quality and risk of bias of research methods, internal validity, reported outcomes, and generalizability.

Statistical Analysis

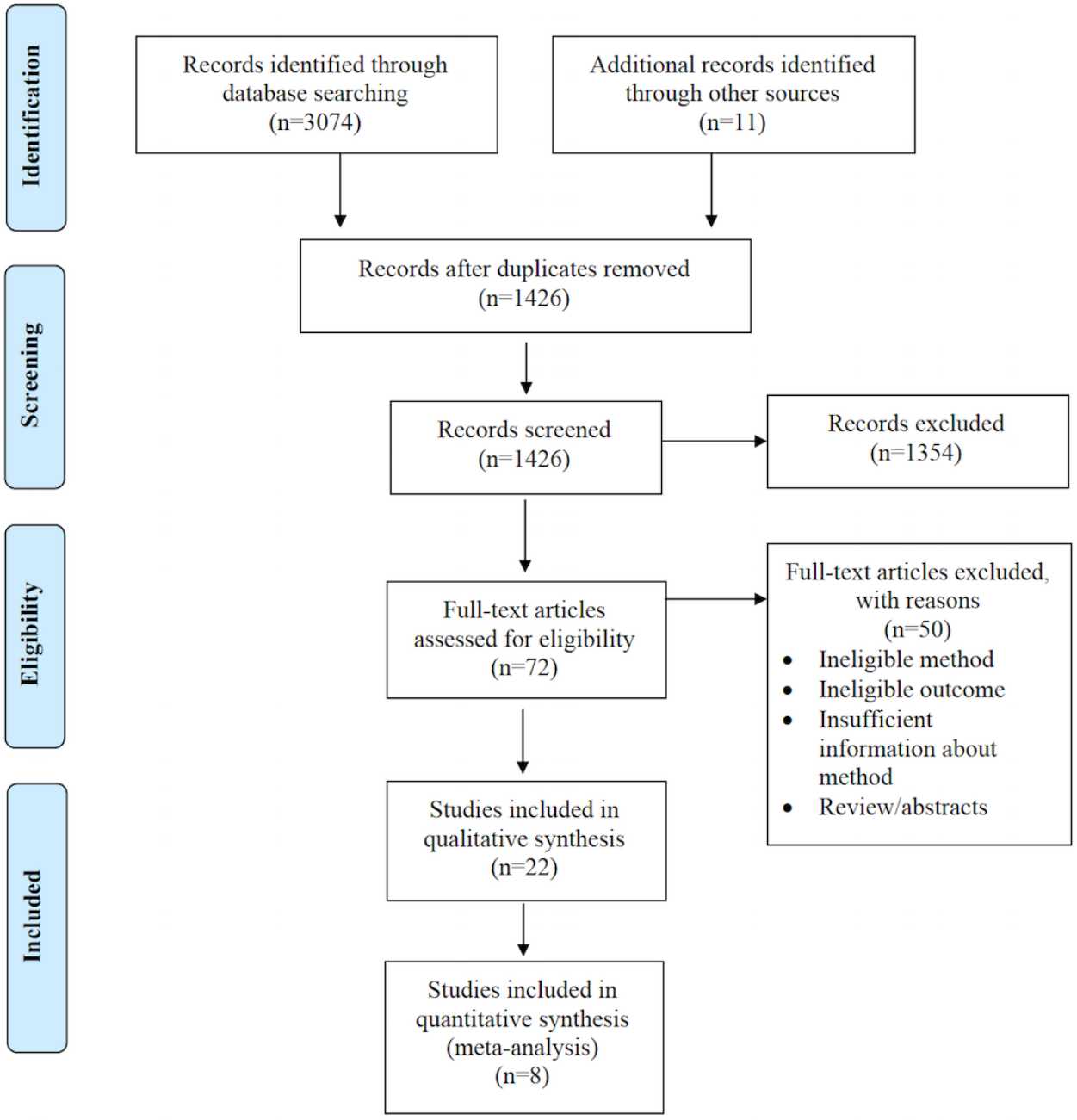
For studies that compared Fitbit with PSG and provided quantitative data, raw Hedges g effect sizes of SOL, WASO, TST, and SE were calculated as the mean differences between average values provided by Fitbit and PSG divided by the standard deviation of PSG values multiplied by the Hedges correction factor [16]. A positive effect size infers lower values derived by Fitbit relative to those derived by PSG. Overall effect size and 95% prediction interval for each parameter per nonsleep-staging and sleep-staging Fitbit models were calculated using a random-effects model [17,18]. Forest plots were created by Microsoft Excel for Mac, version 16.25 (Microsoft Corporation). A threshold probability of 5% (P=.05) was selected as the basis for rejecting the null hypothesis, effect size equals zero. Effect size values of 0.2, 0.5, and 0.8 are considered small, medium, and large effects, respectively [19]. The null hypothesis stating that studies share a common effect size per sleep parameter was tested by calculating the Q statistic [17]. τ^2 represents the overall variation of true effect size, and I^2 represents the proportion of observed variance indicative of actual variation among studies [17]. I^2 values in the order of 25%, 50%, and 75% are considered small, medium, and large heterogenicity, respectively [20]. Comparisons between nonsleep-staging and sleep-staging Fitbit models were accomplished by random-effect subgroup analyses. As recommended, the threshold probability of 10% (P=.10) was the basis for testing the significance of heterogenicity and also for determining statistical significance of subgroup comparisons [20,21].

Results

Search Results

Figure 1 presents a visual summary of the selection and qualification of articles for review. A total of 3074 publications were retrieved though a search of databases performed in July 2018 and again in July and October 2019. An additional 11 publications were identified through other sources, primarily through the reference list of identified articles. After eliminating duplicate publications found in the multiple databases, 1426 articles remained for screening. Examination of individual titles and abstracts yielded 72 publications for full-text appraisal; however, after scrutinizing each of them according to the a priori inclusion and exclusion criteria, only 22 qualified for systematic review [22-43], with 8 of these reporting raw data to enable quantitative synthesis and meta-analyses.

Figure 1. Flow diagram adapted from Moher et al [14] describing the search strategy of databases to retrieve and qualify publications of relevance for review.



Overview of Included Studies

Tables 1 and 2 present the extracted details of each qualifying study involving nonsleep-staging and sleep-staging Fitbit models. Participants were diverse: normal sleepers as well as persons diagnosed with periodic limb movement in sleep (PLMS) [28], obstructive sleep apnea, sleep-disordered breathing [30], central disorders of hypersomnolence [26], insomnia [31], and depression [25], and Huntington disease gene carriers [38]. Sample size varied substantially between investigations, from 7 to 63 (median 30) participants, with approximately 77% of them involving more than 20 individuals. Average age was less than 20 years in 6 of the 22 studies (27%) and over 50 years in 1 study (5%). Out of 22 studies, 10 (45%) were conducted in a sleep laboratory, 11 (50%) in the home environment, and 1 (5%) either at home or in a hotel, based on the participant's preference.

Table 1. Detailed summary of qualifying publications involving both early-generation, nonsleep-staging and newer-generation, sleep-staging Fitbit models.

Author (year)	Model	Reference	Participants			Investigative details		
			Total number (% female)	Age (years), mean (SD); and/or range	Type	Duration (study site)	Tracker placement (tracker mode)	Bedtime
Beattie et al (2017) [22][a]	Surge	Type III home PSG[b]	60 (40.0)	34 (10)	Normal sleepers	1 night (home or hotel)	Both left and right wrists (N/A[c])	22:00h
Brazendale et al (2019) [23]	Charge HR	Sleep log & actigraphy	30 (37.0)	7.2 (2.1)	Healthy	2 nights (home)	Nondominant wrist (N/R[d])	Habitual[e]
Brooke et al (2017) [24]	Flex & Charge HR	Sleep log	95 (64.2): 22 Flex; 14 Charge HR	28.5 (9.9); 19-60	Healthy	1 night (home)	Left wrist (N/R)	Habitual[e]
Cook et al (2017) [25]	Flex	PSG & actigraphy	21 (81.0)	26.5 (4.6)	Major depressive disorder	1 night (sleep lab)	Nondominant wrist (both sensitive and normal)	Habitual
Cook et al (2019) [26][a]	Alta HR	PSG	49 (93.9)	30.3 (9.8)	Suspected CDH[f]	1 night (sleep lab)	Nondominant wrist (N/A)	Participant preference
de Zambotti et al (2016) [27]	Charge HR	PSG	32 (46.9)	17.3 (2.5); 12-21	Healthy adolescents	1 night (sleep lab)	Nondominant wrist (normal)	Participant preference
de Zambotti et al (2018) [28][a]	Charge 2	PSG	44 (59.1)	19-61	Healthy adults, 9 with PLMS[g]	1 night (sleep lab)	Nondominant wrist (N/A)	Participant preference
Dickinson et al (2016) [29]	Charge HR	Actigraphy	38 (60.5)	26.1 (8.0)	Young adults	4 nights (home)	N/R (normal)	Habitual[e]
Hakim et al (2018) [30]	Charge HR	PSG	22 (59.1)	9 (3); 3-18	OSA[h] or SDB[i]	1 night (sleep lab)	N/R (N/R)	N/R
Kang et al (2017) [31]	Flex	Unattended PSG	Insomniacs: 33 (57.6); good sleepers: 17 (64.7)	Insomniacs: 38.4 (11.2); good sleepers: 32.1 (7.4)	Insomniacs and good sleepers	1 night (home)	Nondominant wrist (both sensitive and normal)	N/R
Kubala et al (2019) [32]	Alta	Actigraphy	Good sleepers: 20 (60); poor sleepers 10 (30)	24.8 (4.1); 18-60	Healthy adults	7 nights: Fitbit was worn for only 1 night (home)	Nondominant wrist (normal)	Habitual[e]
Lee et al (2017) [33]	Charge HR	Actigraphy	16 (62.5)	22.8 (2.8); 18-26	Healthy young adults	13 nights (home)	Nondominant hand (normal)	Habitual[e]
Lee et al (2018) [34]	Charge HR	Sleep log	38 (50)	28.6 (N/R); 19-66	Healthy	3 nights (home)	Counter-balanced across participants (N/R)	Habitual[e]
Liang and Chapa Martell (2018) [35]	Charge 2	Sleep Scope (EEG[j] based)	25 (40.0)	24.8 (4.4)	Healthy	3 nights: only 1 night in data analysis (home)	Nondominant hand (normal)	Habitual[e]
Liu et al (2019) [36][a]	Alta HR	Sleep log	10 (50.0)	21.8 (N/R); 20-24	Noninsomniac Asian students	7 nights (home)	Left wrist (N/A)	Habitual[e]

Author (year)	Model	Reference	Participants			Investigative details		
			Total number (% female)	Age (years), mean (SD); and/or range	Type	Duration (study site)	Tracker placement (tracker mode)	Bedtime
Mantua et al (2016) [37]	Flex	PSG (ambulatory system)	40 (47.5)	22.4 (4.9); 18-30	Healthy young adults	1 night (home)	Counter-balanced across participants (N/R)	N/R
Maskevich et al (2017) [38]	One	PSG	7 (85.7)	54.1 (6.4)	Huntington disease gene carriers	1 night (sleep lab)	Nondominant wrist (default)	Habitual
Meltzer et al (2015) [39]	Ultra	PSG & actigraphy (2 different ones)	63 (50.8)	9.7 (4.6); 3-17	Children and adolescents	1 night (sleep lab)	Nondominant wrist (both sensitive and normal)	Habitual
Montgomery-Downs et al (2012) [40]	Classic	PSG & actigraphy	24 (40)	26.1 (N/R); 19-41	Healthy adults	1 night (sleep lab)	Nondominant wrist (N/R)	N/R
Osterbauer et al (2016) [41]	Flex	PSG	14 (64.3)	6.5 (2.9); 3-11	Children	1 night (sleep lab)	Nondominant wrist (N/R)	N/R
Sargent et al (2018) [42]	Charge HR	PSG	12 (N/R)	18.3 (1.0)	Soccer players without sleep disorders	5 sleep periods during 3 days and nights (sleep lab)	Nondominant wrist (normal)	22:00-07:00h); 23:00-07:00h); 00:00-07:00h); 14:00-16:00h); 15:00-16:00h
Svensson et al (2019) [43][a]	Versa	Sleep Scope (EEG based)	20 (50)	25-67	Healthy Japanese adults	14 nights: only 7 nights in data analysis (home)	Nondominant wrist (normal)	Habitual[e]

[a]Publication consisting of newer-generation, sleep-staging Fitbit models.

[b]PSG: polysomnography.

[c]N/A: not applicable.

[d]N/R: not reported.

[e]Exact bedtime not reported, but investigative protocol infers habitual bedtime.

[f]CDH: central disorders of hypersomnolence.

[g]PLMS: periodic limb movement in sleep.

[h]OSA: obstructive sleep apnea.

[i]SDB: sleep-disordered breathing.

[j]EEG: electroencephalographic.

Table 2. Results from qualifying publications involving both early-generation, nonsleep-staging and newer-generation, sleep-staging Fitbit models.

Author (year)	Model	Reference	Results[a]
Beattie et al (2017) [22][b]	Surge	Type III home PSG[c]	• Normal mode vs PSG. Overestimated TST[d,e] (46 min) and SE[e,g] (8.1%); underestimated WASO[e,h] (44 min) and SOL[i] (2 min, NS[j]); accuracy, 0.88 (SD 0.05); sensitivity, 0.98 (SD 0.02); specificity, 0.35 (SD 0.13) • Normal mode vs actigraphy. Overestimated SOL[e] (12 min), SE[f] (1.1%), and TST (5 min, NS); underestimated WASO[e] (17 min) • Sensitive mode vs PSG. Underestimated TST[e] (86 min) and SE[e] (16.0%); overestimated SOL[f] (12 min) and WASO[e] (75 min); accuracy, 0.78 (SD 0.08); sensitivity, 0.78 (SD 0.09); specificity, 0.80 (SD 0.17) • Sensitive mode vs actigraphy. Underestimated TST[e] (127 min) and SE[e] (22.9%); overestimated SOL[e] (25 min) and WASO[e] (102 min)
Brazendale et al (2019) [23]	Charge HR	Sleep log & actigraphy	• Fitbit correlated with both actigraphy (r=.48)[f] and sleep log (r=.71)[f] for measuring TST
Brooke et al (2017) [24]	Flex & Charge HR	Sleep log	• Both Fitbit Flex (r=.68, MAPE[k]=8.80%) and Fitbit Charge HR (r=.58, MAPE=11.5%) correlated with sleep log in measuring TST[e]
Cook et al (2017) [25]	Flex	PSG & actigraphy	• Normal mode vs PSG. Overestimated TST[e] (46 min) and SE[e] (8.1%); underestimated WASO[e] (44 min) and SOL (2 min, NS); accuracy, 0.88 (SD 0.05); sensitivity, 0.98 (SD 0.02); specificity, 0.35 (SD 0.13) • Normal mode vs actigraphy. Overestimated SOL[e] (12 min), SE[f] (1.1%), and TST (5 min, NS); underestimated WASO[e] (17 min) • Sensitive mode vs PSG. Underestimated TST[e] (86 min) and SE[e] (16.0%); overestimated SOL[f] (12 min) and WASO[e] (75 min); accuracy, 0.78 (SD 0.08); sensitivity, 0.78 (SD 0.09); specificity, 0.80 (SD 0.17) • Sensitive mode vs actigraphy. Underestimated TST[e] (127 min) and SE[e] (22.9%); overestimated SOL[e] (25 min) and WASO[e] (102 min)
Cook et al (2019) [26][b]	Alta HR	PSG	• Overestimated TST[f] (12 min), SE[f] (2.0%), and deep sleep[e] (18 min); underestimated SOL (4 min, NS), WASO (8 min, NS), and light sleep (11 min, NS); accuracy, 0.90 (SD 0.04); sensitivity, 0.96 (SD 0.02); specificity, 0.58 (SD 0.16); accuracy in detecting light sleep, 0.73; deep sleep, 0.89; REM[l] sleep, 0.89
de Zambotti et al (2016) [27]	Charge HR	PSG	• Overestimated TST[f] (8 min) and SE[e] (1.8%); underestimated WASO[f] (6 min) and SOL (3 min, NS); accuracy, 0.91 (SD 0.05); sensitivity, 0.97 (SD 0.02); specificity, 0.42 (SD 0.16); predictive value for sleep, 0.93 (SD 0.05); predictive value for wake, 0.65 (SD 0.18)
de Zambotti et al (2018) [28][b]	Charge 2	PSG	• Normal sleeper cohort. Overestimated TST[f] (9 min) and light sleep[e] (34 min); underestimated SOL[f] (4 min), deep sleep[e] (24 min), WASO (5 min, NS), and REM sleep (1 min, NS); sensitivity, 0.96; specificity, 0.61; accuracy in detecting light sleep, 0.81; deep sleep, 0.49; REM sleep, 0.74 • PLMS[j] cohort. Underestimated deep sleep[f] (28 min), SOL (7 min, NS), and WASO (1 min, NS); overestimated TST (8 min, NS), light sleep (36 min, NS), and REM sleep (0 min, NS); specificity, 0.62; accuracy detecting light sleep, 0.78; deep sleep, 0.36; REM sleep, 0.62
Dickinson et al (2016) [29]	Charge HR	Actigraphy	• No systematic difference across days between Fitbit and actigraphy in measuring TST and SE
Hakim et al (2018) [30]	Charge HR	PSG	• Overestimated TST[f] (30 min); underestimated total wake time[f] (23 min)
Kang et al (2017) [31]	Flex	Unattended PSG	• Good sleepers—normal mode. Overestimated TST[f] (7 min), SE (1.8%, NS), and SOL (1 min, NS); underestimated WASO (7 min, NS); accuracy, 0.93; sensitivity, 0.97; specificity, 0.36 • Insomniacs—normal mode. Overestimated TST[e] (33 min) and SE[e] (7.9%); underestimated WASO[e] (31 min) and SOL (2.4%, NS); accuracy, 0.87; sensitivity, 0.97; specificity, 0.36 • Good sleepers—sensitive mode. Accuracy, 0.66; sensitivity, 0.65; specificity, 0.82 • Insomniacs—sensitive mode. Accuracy, 0.68; sensitivity, 0.64; specificity, 0.89
Kubala et al (2019) [32]	Alta	Actigraphy	• Good sleepers. Overestimated TST[e] (74 min); underestimated WASO[f] (16 min) • Poor sleepers. Overestimated TST (20 min, NS); underestimated WASO (13 min, NS)

Author (year)	Model	Reference	Results[a]
Lee et al (2017) [33]	Charge HR	Actigraphy	• Overestimated TST[e] (22 min); correlation between Fitbit and actigraphy: sleep start times[e] (r=.87) and TST[e] (r=.92)
Lee et al (2018) [34]	Charge HR	Sleep log	• Correlation between Fitbit and sleep log: TST[e] (r=.55, MAPE 14.2%) and TIB[e,n] (r=.48, MAPE 12.7%); SE and WASO not correlated
Liang and Chapa Martell (2018) [35]	Charge 2	Sleep Scope (EEG[o] based)	• Overestimated WASO[e] (25 min) and deep sleep[e] (40 min); underestimated TST[f] (12 min), SOL[e] (11 min), REM[e] sleep (12 min), light sleep[e] (42 min), and SE (1.5%, NS)
Liu et al (2019) [36][b]	Alta HR	Sleep log	• Overestimated WASO[f] (13 min); underestimated TST[f] (6 min), SOL[f] (5 min), and SE[f] (1.4%)
Mantua et al (2016) [37]	Flex	PSG (ambulatory system)	• No significant difference in measuring TST and SE; TST correlated[e] (r=.97); SE not correlated (r=.21, NS); average percentage error: TST, 2.97%; SE, 11.57%
Maskevich et al (2017) [38]	One	PSG	• Overestimated TST[e] (88 min) and SE[e] (17.4%); underestimated WASO[f] (39 min) and SOL (17 min, NS); accuracy, 0.81 (0.68-0.93); sensitivity, 0.99 (0.97-1.00); specificity, 0.27 (0.12-0.55); predictive value for sleep, 0.99, and wake, 0.27
Meltzer et al (2015) [39]	Ultra	PSG & actigraphy (2 different ones)	• Fitbit—normal mode vs PSG. Underestimated WASO[e] (32 min); overestimated TST[e] (41 min) and SE[e] (8%); accuracy, 0.84; sensitivity, 0.87; specificity, 0.52 • Fitbit—sensitive mode vs PSG. Underestimated TST[e] (105 min) and SE[e] (21%); overestimated WASO[e] (106 min); accuracy, 0.71; sensitivity, 0.70; specificity, 0.79
Montgomery-Downs et al (2012) [40]	Classic	PSG & actigraphy	• Fitbit vs PSG. Overestimated SE[e] (14.5%) and TST[e] (67 min); sensitivity, 0.98 (0.92-1.00); specificity, 0.20 (0.02-0.78) • Fitbit vs actigraphy. Overestimated SE[e] (5.2%) and TST[e] (24 min)
Osterbauer et al (2016) [41]	Flex	PSG	• TST by Fitbit and PSG correlated[f] (rho[p]=.99); WASO, SE, and awake minutes not correlated; sensitivity, 0.99; specificity, 0.10
Sargent et al (2018) [42]	Charge HR	PSG	• TST by Fitbit vs PSG: NS; Fitbit automatically identified 60% of sleep periods, with a success rate of 80% when sleep was 9h, 90% when sleep was 8h, 70% when sleep was 7h, 50% when sleep was 2h, and 10% when sleep was 1h
Svensson et al (2019) [43][b]	Versa	Sleep Scope (EEG based)	• Overestimated TIB (9 min, NS), TST (7 min, NS), WASO[e] (14 min), and deep sleep[e] (36 min); underestimated SE (0.1%, NS), SOL[e] (14 min), REM[e] sleep (6 min), and light sleep[e] (20 min); accuracy, 0.89 (0.88-0.89); sensitivity, 0.92 (0.919-0.923); specificity, 0.54 (0.53-0.55)

[a]Accuracy, sensitivity, and specificity in detecting sleep epochs are reported unless otherwise specified.

[b]Publication consisting of newer-generation, sleep-staging Fitbit models.

[c]PSG: polysomnography.

[d]TST: total sleep time.

[e]$P<.01$.

[f]$P<.05$.

[g]SE: sleep efficiency.

[h]WASO: wake after sleep onset.

[i]SOL: sleep onset latency.

[j]NS: not significant.

[k]MAPE: mean absolute percent error.

[l]REM: rapid eye movement.

[m]PLMS: periodic limb movement in sleep.

[n]TIB: time in bed.

[o]EEG: electroencephalographic.

[p]Spearman correlation coefficient.

Bias Assessment

Table S1 in Multimedia Appendix 1 summarizes the risk for bias of research methods, internal validity, reported outcomes, and generalizability of each qualifying study. Only 5 investigations out of 22 (23%) attempted to blind the sleep laboratory technicians from outcome measures [22,25,26,28,40]. A total of 13 of the qualifying 22 studies (59%) relied on PSG as reference to evaluate Fitbit performance [22,25-28,30,31,37-42]; the other 9 (41%) relied on a sleep log, actigraphy, or home EEG as reference. Since PSG is the *gold standard* for measurement of sleep stages and parameters, use of other methods of reference constitutes an additional potential source of bias. Half of the investigations (11/22, 50%) were performed in the participant's home. Clock hour of bedtime was unspecified in 5 of the 22 studies (23%), and the dictated bedtime differed from the habitual one of participants in 2 laboratory studies (9%). Disparity between habitual versus mandated bedtime might have resulted in a greater- or lesser-than-usual amount of wake time while in bed and this is likely to have led to bias. Moreover, use of somewhat bulky sleep trackers as reference for studies conducted in one's residence might be problematic because of improper donning of instrumentation and oversight by technicians. Most studies (16/22, 73%) followed the manufacturer's recommendation that the Fitbit model be worn on the nondominant hand. However, 2 out of the 22 studies (9%) utilized a counterbalance experimental design for placement of the Fitbit device, with half of the participants wearing it on the right hand and the other half wearing it on the left hand. Finally, in 2 out of the 22 studies (9%), the Fitbit was worn on the left wrist; in 2 other studies (9%), the hand upon which the Fitbit was worn was unreported. Out of the 22 studies, 1 (5%) [28] did not report the SE for Fitbit but did report it for PSG; this may constitute selective reporting bias. We contacted the corresponding author of this investigation for the missing information, but we did not obtain a response.

Comparison of Sleep Parameters Assessed by Fitbit Versus Polysomnography

Nonsleep-Staging Fitbit Models

Out of the 22 studies, 10 (45%) assessed early-generation nonsleep-staging Fitbit models in comparison with PSG in estimating sleep parameters [25,27,30,31,37-42]; 1 of these studies involved performance of Fitbit models when applied to individuals of two different cohorts (ie, good sleepers and insomniacs), thereby increasing the number of possible comparisons to 11. Eight (N=203) of the 10 potential comparisons reported significant overestimation of TST by Fitbit versus PSG of between 6.5 and 88.1 minutes, while the two others (N=52) found nonsignificant overestimation. Five (N=142) of the six potential comparisons reported significant underestimation of WASO by Fitbit versus PSG of between 5.6 and 44 minutes, while one other (N=17) reported nonsignificant underestimation. Six (N=166) of the eight potential comparisons observed significant overestimation of SE by Fitbit versus PSG of between 1.8% and 17.4%, while two others (N=57) reported nonsignificant overestimation. A total of 5 comparisons (N=110) evaluated SOL, finding no significant difference between the two methods of appraisal.

Figures 2-5 present the forest plots of effect size plus the overall pooled estimated effect size of the SOL, WASO, TST, and SE variables derived by nonsleep-staging Fitbit models. The pooled estimate of effect size reveals the following by nonsleep-staging Fitbit models relative to PSG: nonsignificant difference in estimation of SOL (N=4 comparisons; effect size=0.12, 95% CI -0.14 to 0.39; P=.37); significant underestimation of WASO (N=5 comparisons; effect size=0.60, 95% CI 0.38 to 0.83; P<.001); significant overestimation of TST (N=7 comparisons; effect size=-0.51, 95% CI -0.71 to -0.30; P<.001); and significant overestimation of SE (N=6 comparisons; effect size=-0.74, 95% CI -0.97 to -0.48; P<.001). Heterogeneity was not detected in any sleep parameter.

Sleep-Staging Fitbit Models

Only 3 publications out of 22 (14%) pertained to recent-generation sleep-staging Fitbit models in comparison with PSG as reference [22,26,28]; 1 of these 3 publications studied the performance of Fitbit on two different cohorts—normal and PLMS sleepers—thereby increasing the number of possible comparisons to four.

Two (N=84) of the three potential comparisons reported significant overestimation of TST by Fitbit versus PSG in the amount of 9-11.6 minutes, while one other (N=9) reported nonsignificant overestimation. The only study (N=49) that assessed SE reported 1.98% significant overestimation by Fitbit relative to PSG. A total of 3 trials (N=93) reported no significant difference in WASO between the two methods; 1 of these (N=35) found significant 4-minute underestimation of SOL by Fitbit versus PSG, while the 2 others (N=58) detected nonsignificant underestimation of SOL by Fitbit.

The pooled estimate of effect size (see Figures 2-5) revealed a significant underestimation of SOL (N=3 comparisons; effect size=0.32, 95% CI 0.04 to 0.60; P=.03) and a nonsignificant difference in estimation of WASO (N=3 comparisons; effect size=0.16, 95% CI -0.12 to 0.44; P=.25), TST (N=3 comparisons; effect size=-0.15, 95% CI -0.43 to 0.13; P=.29), and SE (N=1 comparison; effect size=-0.27, 95% CI -0.65 to 0.13; P=.19) by sleep-staging Fitbit models versus PSG. Heterogeneity was not detected in any sleep parameter. Since only 1 study evaluated SE, testing for heterogeneity was not relevant.

Figure 2. Forest plot of the standardized mean difference (Hedges g) between Fitbit and polysomnography for the variable of sleep onset latency (SOL). Results are shown as effect size (ES) and 95% CI. The difference in symbol size indicates the difference in weight of the respective studies. The diamond symbol shows the 95% CI of the overall effect and the tails show the 95% prediction interval of the overall effect. PLMS: periodic limb movement in sleep.

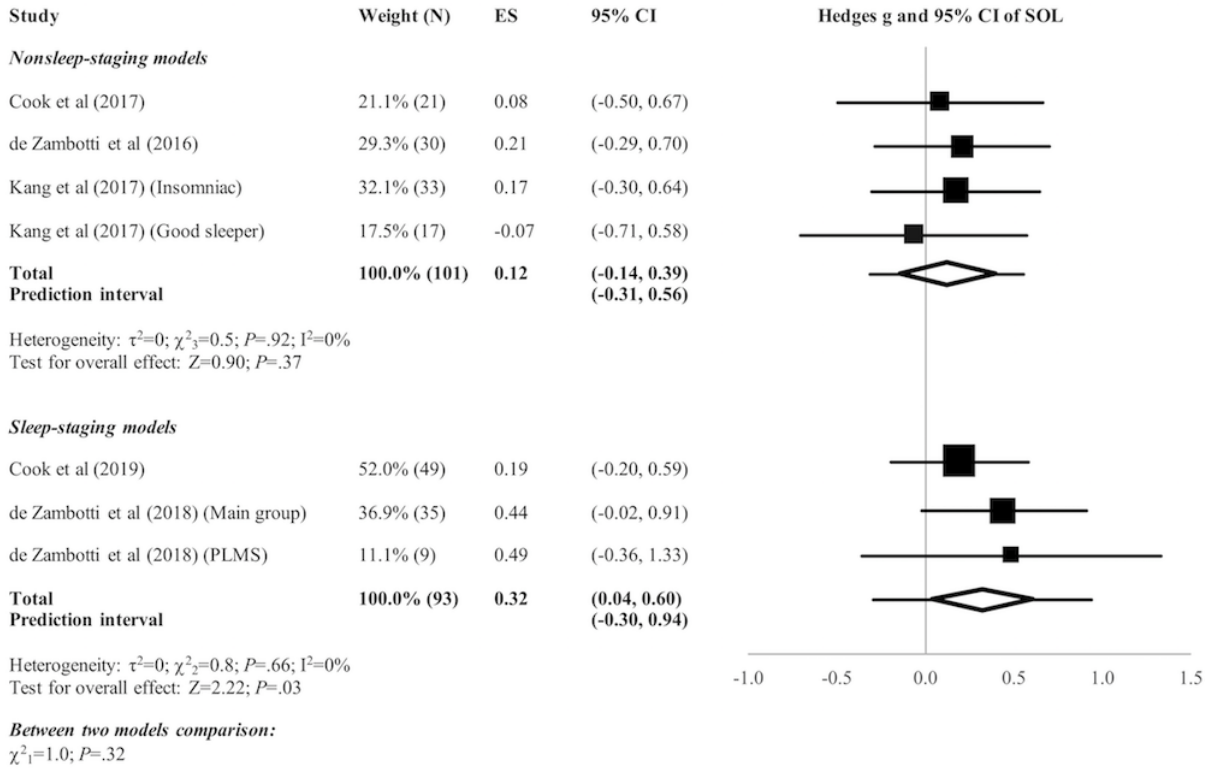


Figure 3. Forest plot of the standardized mean difference (Hedges g) between Fitbit and polysomnography for the variable of wake after sleep onset (WASO). Results are shown as effect size (ES) and 95% CI. The difference in symbol size indicates the difference in the weight of the respective studies. The diamond symbol shows the 95% CI of the overall effect and the tails show the 95% prediction interval of the overall effect. PLMS: periodic limb movement in sleep.

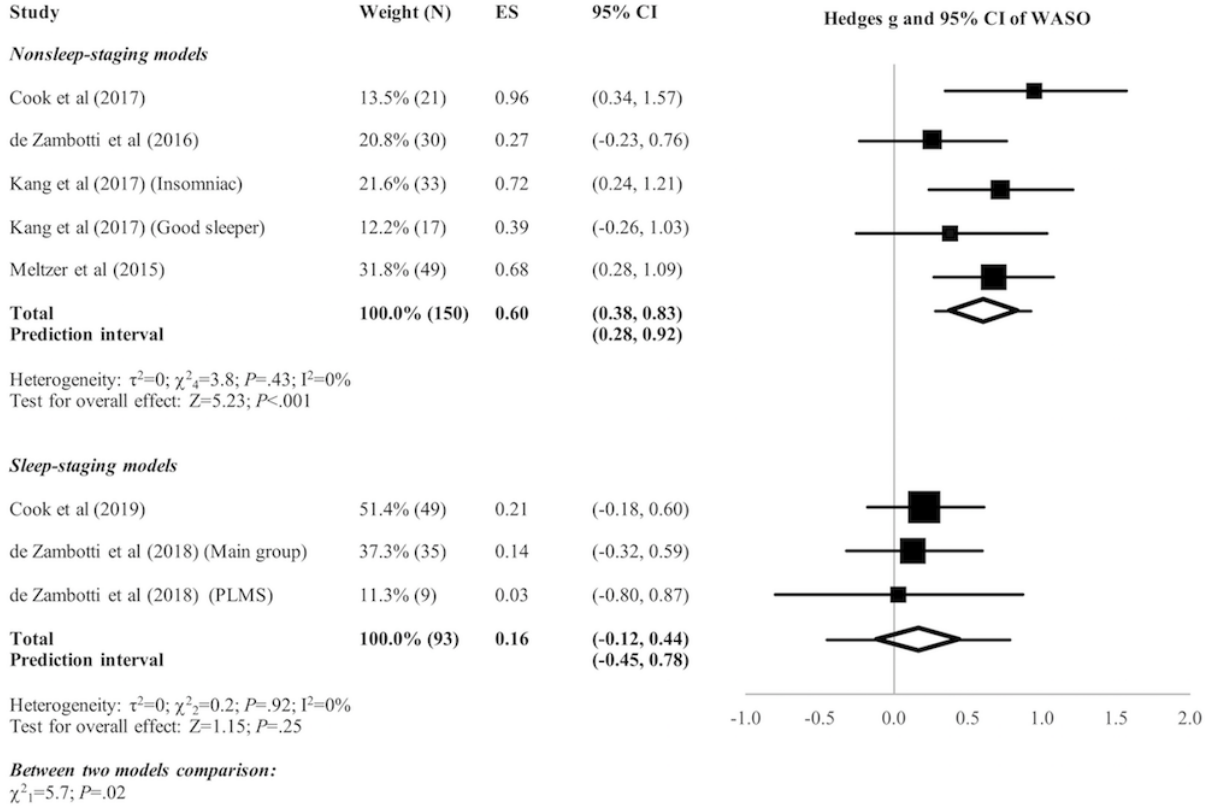


Figure 4. Forest plot of the standardized mean difference (Hedges g) between Fitbit and polysomnography for the variable of total sleep time (TST). Results are shown as effect size (ES) and 95% CI. The difference in symbol size indicates the difference in weight of the respective studies. The diamond symbol shows the 95% CI of the overall effect and the tails show the 95% prediction interval of the overall effect. PLMS: periodic limb movement in sleep.

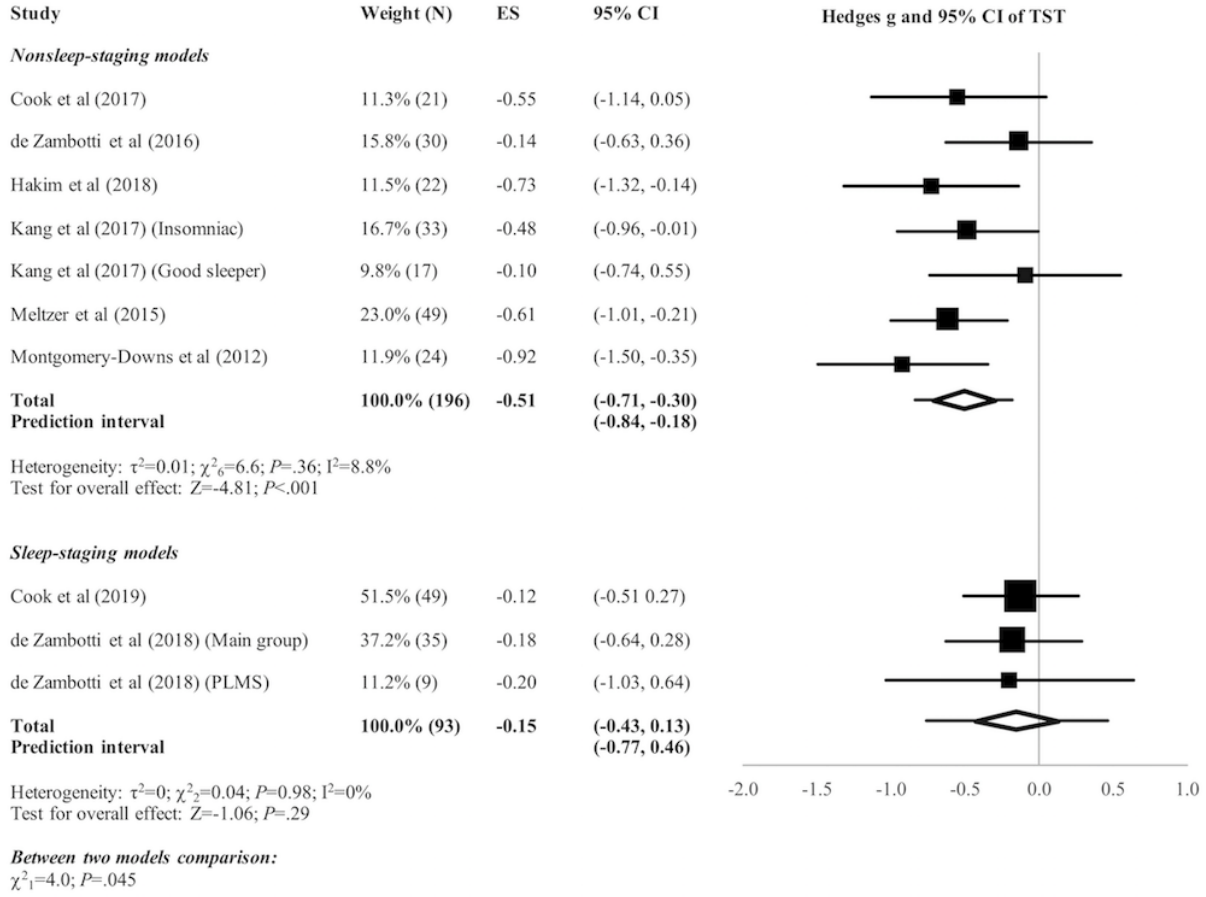
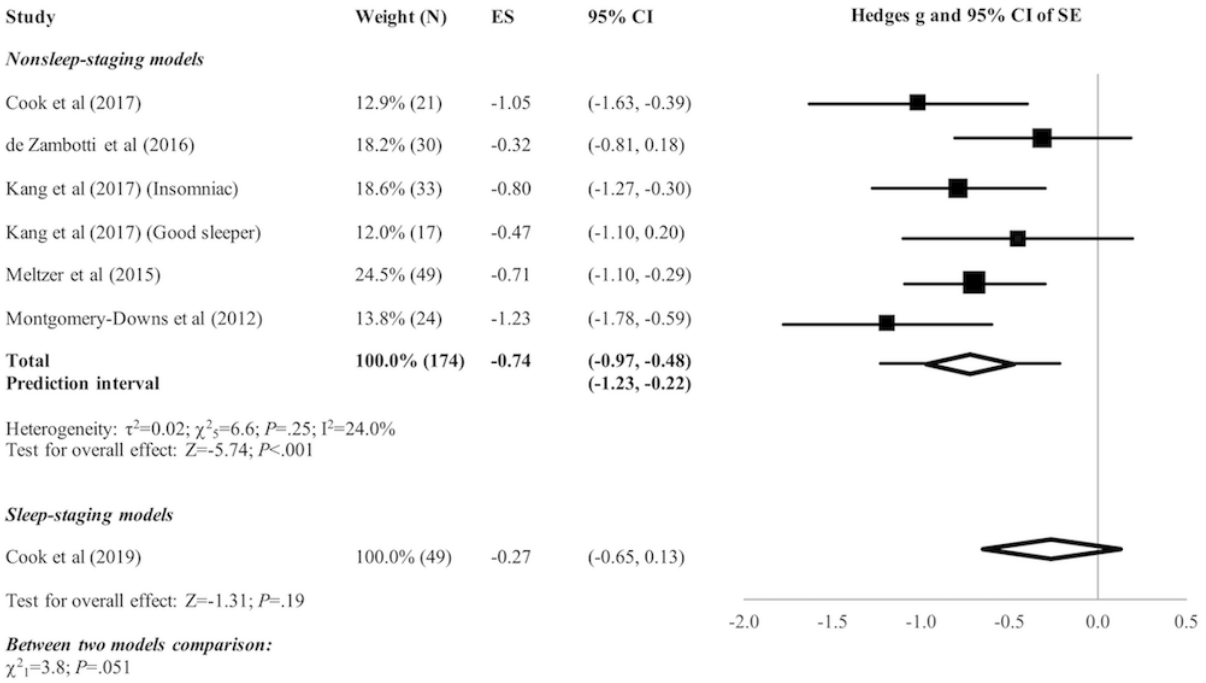


Figure 5. Forest plot of the standardized mean difference (Hedges g) between Fitbit and polysomnography for the variable of sleep efficiency (SE). Results are shown as effect size (ES) and 95% CI. The difference in symbol size indicates the difference in weight of the respective studies. The diamond symbol shows the 95% CI of the overall effect and the tails show the 95% prediction interval of the overall effect.



Comparison of Nonsleep-Staging Versus Sleep-Staging Fitbit Models

Subgroup analyses revealed no significant difference between nonsleep-staging and sleep-staging models in estimating SOL (χ^2_1=1.0, *P*=.32). However, sleep-staging models performed better than nonsleep-staging models in estimating WASO (χ^2_1=5.7, *P*=.02), TST (χ^2_1=4.0, *P*=.045), and SE (χ^2_1=3.8, *P*=.051, ie, below the cutoff of *P*=.10 recommended for statistical significance of subgroup comparisons).

Accuracy, Sensitivity, and Specificity in Detecting Sleep Epochs by Fitbit Versus Polysomnography

Nonsleep-Staging Fitbit Models

A total of 7 studies (N=197) involved epoch-by-epoch (EBE) investigation of nonsleep-staging Fitbit models with reference to PSG in differentiating between sleep and wake state [25,27,31,38-41]. Out of these 7 studies, 1 consisted of two different samples, thereby increasing the total number of evaluations to 8. Across these trials, Fitbit versus PSG analyses identified sleep epochs with accuracy values between 0.81 and 0.93, sensitivity values between 0.87 and 0.99, and specificity values between 0.10 and 0.52.

Sleep-Staging Fitbit Models

Sleep Epoch Identification

A total of 3 studies (N=153) evaluated the performance of sleep-staging Fitbit models with PSG as reference in identifying sleep epochs through EBE analyses [22,26,28]. A total of 1 study included two different samples, thereby increasing the number of possible comparisons to four. Relative to PSG, detection of sleep epochs in three possible comparisons (N=144) revealed sensitivity values between 0.95 and 0.96; detection of sleep epochs in four comparisons (N=153) revealed specificity values between 0.58 and 0.69; and detection of sleep epochs, assessed in only a single study (N=49), revealed an accuracy of 0.90.

Sleep-Stage Identification

A total of 3 studies (N=153) appraised performance of sleep-staging Fitbit models in identifying sleep stages by means of EBE analysis [22,26,28]. Relative to PSG, accuracy varied between 0.69 and 0.81 in detecting light sleep (non-rapid eye movement [REM] 1 [N1] + non-REM 2 [N2]), between 0.36 and 0.89 in detecting deep sleep (non-REM 3 [N3]), and between 0.62 and 0.89 in detecting REM sleep.

Evaluation of Sleep Parameters Estimated by Fitbit Models Versus Actigraphy

No published study assessed performance of sleep-staging Fitbit models relative to actigraphy in estimating sleep parameters. In contrast, 7 studies investigated the accuracy of nonsleep-staging Fitbit models relative to actigraphy [23,25,29,32,33,39,40]; 1 of them involved two different actigraph devices [39], and another 1 included two different samples: good sleepers and poor sleepers [32]. Among the total of seven potential comparisons, Fitbit significantly overestimated TST in five of them (N=84) by 24.1-74 minutes, while in two comparisons (N=31), it was nonsignificantly overestimated. In a total of four comparisons (N=69), Fitbit significantly overestimated SE by 1.1%-7.0%. Among a total of five comparisons, Fitbit significantly underestimated WASO in four of them (N=65) by 16-32 minutes, while in one (N=10), it was nonsignificantly underestimated. Only 1 study (N=21) evaluated SOL, finding an 11.5-minute significant overestimation by Fitbit.

Two studies [25,38] simultaneously compared in the same cohort a nonsleep-staging Fitbit model plus actigraph against laboratory PSG. Relative to PSG, both the actigraph and Fitbit overestimated TST: Fitbit in one study by 88 minutes and in the other by 46 minutes; actigraph in one study by 74 minutes and in the other by 40.6 minutes. They also overestimated SE: Fitbit in one study by 17.4% and in the other by 8.1%; actigraph in one study by 14.8% and in the other by 7%. However, they underestimated WASO: Fitbit in one study by 39 minutes and in the other by 44 minutes; actigraph in one study by 20 minutes and in the other by 27 minutes. Actigraphy showed less bias than Fitbit. On the other hand, in the same two studies, Fitbit performed better than actigraphy in measuring SOL relative to PSG as reference: Fitbit bias in one study of 17 minutes and 2 minutes in the other; actigraph bias in one study of 23 minutes and 14 minutes in the other. One of these studies also performed EBE analysis [25], finding approximately 1% higher sensitivity and accuracy in detecting sleep by Fitbit than by actigraphy relative to PSG.

Correlation Between Sleep Parameters Assessed by Fitbit Versus Sleep Diary

A total of 3 studies investigated the extent of correlation between sleep parameters derived by a nonsleep-staging Fitbit model and a self-rated sleep diary [23,34]; 1 of them involved two different Fitbit models [24]. In the total of four potential comparisons (N=104), significant correlation, ranging between r=.55 and r=.71, was reported between TST measured by Fitbit versus sleep diary. A total of 1 study (N=38) found significant correlation between the two methods of sleep assessment for time in bed (TIB; r=.48) but poor correlation for WASO (r=.09) and SE (r=-.03). A total of 1 study (N=10) [36] investigated the extent of agreement between the sleep diary and sleep-staging Fitbit methods. It reported significant overestimation of WASO (13 minutes) and underestimation of SOL (5 minutes), TST (6 minutes), and SE (1.4%) by Fitbit in comparison to the sleep diary.

Dominant Versus Nondominant Hand Comparison of Fitbit Sleep-Stage Classification Accuracy

Only 1 study (N=60) [22] explored differences in accuracy of sleep-stage classification when new-generation Fitbit models were simultaneously worn on the dominant and nondominant hand. No between-hand difference was found in estimating sleep stages.

Effect of Selected Sensitivity Mode Setting on Estimation of Sleep Parameters

Early-generation nonsleep-staging, but not later-generation sleep-staging, Fitbit models allow the user to select either normal or sensitive mode to sense body movement to derive sleep

parameters. A total of 3 studies evaluated, relative to PSG, performance of these Fitbit models when set to the sensitive mode [25,31,39]. In 2 of them, the sensitive mode significantly underestimated TST by 86.3 minutes (N=21) and 105 minutes (N=63), respectively, and underestimated SE by 16.0% (N=21) and 21% (N=63), respectively. The only study (N=21) that evaluated SOL found that the sensitive mode significantly overestimated it by 11.5 minutes. Across the 3 studies (N=134), differentiation of sleep from wake epochs by early-generation nonsleep-staging Fitbit models when set to the sensitive mode relative to PSG ranged from 0.66 to 0.78 in accuracy, from 0.64 to 0.78 in sensitivity, and from 0.79 to 0.89 in specificity.

Interdevice Reliability

A total of 3 studies evaluated interdevice reliability of nonsleep-staging Fitbit models. In 1 of them, 3 participants wore two Fitbit Classic bands on the same wrist, finding through EBE comparisons high interdevice reliability (96.5%-99.1%) [40]. Another study with 7 participants who wore two Fitbit Ultra devices on the same wrist also substantiated essentially equivalent between-device performance in estimating TST and SE; findings for other sleep parameters were not reported [39]. Finally, the remaining study involving 10 participants who simultaneously wore two Fitbit Alta devices on the wrist of their nondominant arm found no significant difference between the two devices in measuring WASO, but slight, yet statistically significant, difference (approximately 6 minutes) between them in measuring TST [32].

Discussion

Principal Findings

The quality and uses of personal monitoring technology are rapidly advancing, offering the promise of extensive improvement in medical literacy and health. An area of high interest today to consumers and health professionals is sleep quality because of its recognized importance to daytime cognitive and physical performance. A number of wrist-worn devices enable tracking of sleep parameters and stages. Some of the most popular ones are marketed by Fitbit, Inc; performance of several of its nonsleep-staging and sleep-staging models have been evaluated against laboratory PSG, home sleep trackers, or other methods. The objective of this systematic review was to comprehensively evaluate the worthiness of these consumer wristband devices in assessing sleep.

PSG is regarded as the gold standard for assessment of sleep parameters and stages; in comparison to PSG, nonsleep-staging Fitbit models overestimate TST and SE, underestimate WASO, but determine SOL equally well. Moreover, the amount of bias in estimating TST, SE, and WASO by such wristband models is not negligible. EBE analyses demonstrate, in comparison to laboratory PSG, the high accuracy and sensitivity of this type of Fitbit model in detecting sleep; however, the analyses demonstrate only modest specificity.

In 2007, the American Academy of Sleep Medicine certified wrist actigraphy for at-home evaluation of sleep patterns of both healthy adults and patients with certain suspected sleep disorders [44]. However, several studies have found actigraphy used in conjunction with any one of the four popular interpretative algorithms [2] overestimates sleep duration; although its sensitivity in detecting sleep is high (ie, between 0.87 and 0.99), its specificity is low (ie, between 0.28 and 0.67) [3]. This is the case because actigraphy and its interpretative algorithms tend to score epochs of quiet wakefulness as sleep [45]. Studies that compared nonsleep-staging Fitbit models with actigraphy [23,25,29,32,33,39,40] revealed that Fitbit overestimates both TST and SE and underestimates WASO. Only a single study [25] assessed SOL relative to actigraphy; it found that Fitbit significantly overestimated this sleep parameter, on average, by 12 minutes. The single study [25] that performed EBE analyses on the same participants who were simultaneously outfitted with nonsleep-staging Fitbit and actigraph devices reported only a minor (ie, approximately 1% higher) difference in sensitivity and accuracy by Fitbit. In this regard, nonsleep-staging Fitbit models enable user selection of either normal or sensitive mode to measure body movement to derive sleep parameters. The recommended normal mode only scores *significant* body movements while in bed as awake time. In contrast, the sensitive mode interprets nearly all such movements as awake or restless sleep time [46]. Studies that evaluated Fitbit devices relative to PSG when the devices were set to the sensitive and normal modes reported that the sensitive mode, compared to the normal mode, gave rise to notably higher bias in estimating SOL, SE, WASO, and TST; the sensitive mode also gave rise to lower accuracy and sensitivity but higher specificity. These findings confirm the company's recommendation that the normal mode setting be used in most instances.

Fitbit introduced its sleep-staging feature in 2017, which is now incorporated into the Fitbit Charge 2, Fitbit Charge 3, Fitbit Alta HR, Fitbit Versa, Fitbit Versa 2, Fitbit Blaze, and Fitbit Ionic models. This feature relies on a combined body movement and HRV algorithm to identify and estimate time spent in individual sleep stages [5]. The Fitbit interpretative proprietary sleep-staging algorithm was derived using machine learning methods (ie, linear discriminant classifier) applied to three types of parameters—motion, HRV, and respiratory rate, with the last two calculated from heartbeat data sensed by photoplethysmography—measured during a sleep study of 60 normal sleepers, 18-60 years of age (mean 34, SD 10). These three groups of parameters led to an initial set of 180 features, which, through the method of recursive feature elimination, was reduced to 54 features. Subsequently, heartbeat-derived features were compared to movement-derived features and found to be of approximately equal importance in the overall classification and discrimination of sleep versus wake epochs [22]. Based on discussion with Fitbit, Inc, the same core hardware technology and software algorithm have been incorporated into all sleep-staging Fitbit models since their introduction in 2017, thereby making feasible valid performance comparisons across the published studies.

Thus far, only 3 qualifying studies investigated the performance of the newer-generation Fitbit hardware- and software-coupled technology relative to PSG [22,26,28]. Of these, 1 evaluated the performance of Fitbit for two different—normal sleeper and PLMS—cohorts, thereby increasing the number of possible

comparisons to four, depending on the specific sleep parameters evaluated in individual trials. Although these comparisons showed that recent-generation Fitbit models display only moderate accuracy in detecting sleep stages, they were much better at estimating TST, SE, and WASO than were early-generation, nonsleep-staging Fitbit models. Overall, the amount of bias in estimating the majority of these sleep parameters by new-generation Fitbit models in reference to PSG was clinically negligible: less than 12 minutes in the three comparisons in which TST was determined, 1.98% in the single comparison in which SE was assessed [26], and less than 7 minutes in the three comparisons in which SOL was measured. Meta-analysis of the published data also substantiated the lack of a statistically significant difference between sleep-staging Fitbit models and PSG in measuring WASO, TST, and SE, and with effect sizes of differences in the range of *small*, even for SOL. EBE analyses conducted on the data from the same four comparisons also revealed high sensitivity (ie, between 0.95 and 0.96) and specificity (ie, between 0.58 and 0.69) in detecting sleep. These four comparison trials involving sleep-staging Fitbit models disclosed much higher specificity in detecting sleep than all of the nonsleep-staging Fitbit model comparison trials that reported specificity in the range of 0.10-0.52. Moreover, three of these four comparisons found no significant difference in WASO between methods of assessment. In contrast, five comparisons consisting of nonsleep-staging Fitbit models reported significant underestimation of WASO by 5.6-44 minutes. Collectively, these findings imply the body movement and HRV sensor and software algorithm technology of recent-generation sleep-staging Fitbit models, in comparison to the early-generation ones, more accurately detects wake epochs during intended sleep. These are very promising results, since the major drawback of body activity and movement-based trackers, such as actigraphs, is overestimation of sleep time and poor sensitivity of detecting WASO (ie, poor differentiation of wake from sleep epochs during spans of quiet bedtime activity) [45,47].

A concern about commercially available personal monitors is intra- and interdevice reliability. A total of 3 investigations [32,39,40] addressed this matter and they all reported acceptable consistency. The sole longitudinal (ie, 4 consecutive nights) investigation [29] found no systematic internight difference in TST and SE between Fitbit and actigraphy used as reference. The authors of this study concluded that Fitbit can be a useful device to assess trends in sleep quality, even though absolute values of some sleep parameters might be biased [29]. Of these studies, 1 reported a statistically significant, but nonetheless clinically nonsignificant, interdevice difference (approximately 6 minutes) in measuring TST [32].

Subjective self-report survey and diary methods, even though popular means of assessing sleep due to their ease of use and low cost, are of limited value. Self-report survey approaches depend on recall, which can be biased, especially when not restricted to recent, single point-in-time experiences. Sleep diary methods depend less on memory, but reported information may not be sufficiently accurate because of poor awareness of certain events, such as number or frequency of nighttime awakenings and precise time of falling asleep [48]. Studies that compared Fitbit models with sleep diary methods found a significant correlation between the two approaches for TIB and TST. Because of their simplicity and affordability, sleep survey and diary approaches are generally favored over actigraphy; however, the relatively inexpensive objective method of Fitbit models, along with their ease of use and better estimation of most sleep variables, may render them more appealing.

The findings of this systematic review pertaining to both nonsleep-staging and sleep-staging Fitbit models are based on our recent comprehensive search of databases for relevant published articles, which was repeated three different times during a 15-month span. The included research studies have certain limitations (eg, more than half of them were performed in a laboratory rather than a home setting, and participants were mostly young or middle-aged normal sleepers). Moreover, nonsignificant findings of studies with small sample sizes might be the consequence of insufficient statistical power. Furthermore, only 5 of the 22 qualifying published investigations involved Fitbit sleep-staging models. An additional limitation is a lack of published information as to the extent that advances in Fitbit hardware and/or software technology explain the described disparity in performance between the company's different generations of models in deriving sleep parameters and stages. Based on our dialogue with company representatives, this disparity is due collectively to advances in sensor technology; improved fidelity of data acquisition; and incorporation of heart rate, HRV, and body movement into its sleep-study-validated proprietary interpretative algorithm. In spite of these limitations, findings of this review indicate that the advanced body movement and HRV method of recent-generation Fitbit wristband models seems appropriate to derive suitable estimates of sleep parameters and time spent in sleep. The findings further suggest that such Fitbit models may be useful for conduct of population-based sleep research, which, in the past, typically relied heavily or entirely upon subjective methods.

Conclusions

Fitbit models are marketed to the lay public to allow users to self-derive knowledge about their sleep quality, rather than as a substitute for standard clinical polysomnography. They appear useful for the study of the 24-hour sleep-wake pattern and for the determination of the duration, pattern, and quality of sleep, longitudinally, over many consecutive nights under normal living conditions. In this regard, individuals can benefit from information obtained by wristband trackers to improve sleep hygiene and sleep, itself. In certain cases, primary care and sleep specialists can at least gain superficial perspective on the sleep of patients. In spite of the fact that there has yet to be sufficient evaluation of recent-generation sleep-staging Fitbit models, findings of the few, thus far, published studies imply that their performance in differentiating wake from sleep epochs is better than that reported in the literature for actigraphy.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Bias assessment of included studies.
[DOCX File, 22 KB - jmir_v21i11e16273_app1.docx]

References

1. Hirshkowitz M. SlidePlayer. Assessing sleep wearables and in-bedroom devices: CTA standards work URL: https://slideplayer.com/slide/13756482/ [accessed 2019-06-14]
2. Haghayegh S, Khoshnevis S, Smolensky MH, Diller KR, Castriotta RJ. Performance comparison of different interpretative algorithms utilized to derive sleep parameters from wrist actigraphy data. Chronobiol Int 2019 Dec;36(12):1752-1760. [doi: 10.1080/07420528.2019.1679826] [Medline: 31658822]
3. Van de Water AT, Holmes A, Hurley DA. Objective measurements of sleep for non-laboratory settings as alternatives to polysomnography: A systematic review. J Sleep Res 2011 Mar;20(1 Pt 2):183-200 [FREE Full text] [doi: 10.1111/j.1365-2869.2009.00814.x] [Medline: 20374444]
4. Fitbit. Who we are URL: https://www.fitbit.com/about [accessed 2018-10-03]
5. Fitbit. What should I know about sleep stages? URL: https://help.fitbit.com/articles/en_US/Help_article/2163 [accessed 2019-03-29]
6. Haghayegh S, Khoshnevis S, Smolensky MH, Diller KR. Accuracy of PurePulse photoplethysmography technology of Fitbit Charge 2 for assessment of heart rate during sleep. Chronobiol Int 2019 Jul;36(7):927-933. [doi: 10.1080/07420528.2019.1596947] [Medline: 30990098]
7. Thompson WR. Worldwide Survey of Fitness Trends for 2019. ACSMs Health Fit J 2018;22(6):10-17. [doi: 10.1249/fit.0000000000000438]
8. Smith C. DMR. 2019 Sep 06. 60 interesting Fitbit statistics and facts (December 2018) URL: https://expandedramblings.com/index.php/fitbit-statistics/ [accessed 2019-09-05]
9. Wright SP, Collier SR, Brown TS, Sandberg K. An analysis of how consumer physical activity monitors are used in biomedical research. FASEB J 2017 Apr;31(1 (supplement)):1 [FREE Full text]
10. All of Us Research Program Investigators, Denny JC, Rutter JL, Goldstein DB, Philippakis A, Smoller JW, et al. The "All of Us" Research Program. N Engl J Med 2019 Aug 15;381(7):668-676. [doi: 10.1056/NEJMsr1809937] [Medline: 31412182]
11. National Institutes of Health, All of Us Research Program. 2019 Jan 16. All of Us Research Program expands data collection efforts with Fitbit URL: https://allofus.nih.gov/news-events-and-media/announcements/all-us-research-program-expands-data-collection-efforts-fitbit [accessed 2019-09-06]
12. Sperlich B, Holmberg H. Wearable, yes, but able…?: It is time for evidence-based marketing claims!. Br J Sports Med 2017 Aug;51(16):1240 [FREE Full text] [doi: 10.1136/bjsports-2016-097295] [Medline: 27986762]
13. American National Standards Institute. 2016 Sep. CTA/NSF 2052.1-2016 (ANSI): Definitions and characteristics for wearable sleep monitors URL: https://webstore.ansi.org/standards/ansi/ctansf20522016ansi [accessed 2019-10-22]
14. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
15. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. J Epidemiol Community Health 1998 Jul;52(6):377-384 [FREE Full text] [doi: 10.1136/jech.52.6.377] [Medline: 9764259]
16. Coe R. It's the effect size, stupid: What effect size is and why it is important. In: Proceedings of the Annual Conference of the British Educational Research Association. 2002 Presented at: Annual Conference of the British Educational Research Association; September 12-14, 2002; University of Exeter, UK URL: https://www.leeds.ac.uk/educol/documents/00002182.htm
17. Borenstein M, Hedges L, Higgins JPT, Rothstein HR. Introduction to Meta-Analysis. Chichester, UK: John Wiley & Sons; 2009.
18. IntHout J, Ioannidis JPA, Rovers MM, Goeman JJ. Plea for routinely presenting prediction intervals in meta-analysis. BMJ Open 2016 Jul 12;6(7):e010247 [FREE Full text] [doi: 10.1136/bmjopen-2015-010247] [Medline: 27406637]
19. Sullivan GM, Feinn R. Using effect size: Or why the P value is not enough. J Grad Med Educ 2012 Oct;4(3):279-282 [FREE Full text] [doi: 10.4300/JGME-D-12-00156.1] [Medline: 23997866]
20. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003 Oct 06;327(7414):557-560 [FREE Full text] [doi: 10.1136/bmj.327.7414.557] [Medline: 12958120]
21. Richardson M, Garner P, Donegan S. Interpretation of subgroup analyses in systematic reviews: A tutorial. Clin Epidemiol Glob Health 2019 Jun;7(2):192-198. [doi: 10.1016/j.cegh.2018.05.005]

22. Beattie Z, Oyang Y, Statan A, Ghoreyshi A, Pantelopoulos A, Russell A, et al. Estimation of sleep stages in a healthy adult population from optical plethysmography and accelerometer signals. Physiol Meas 2017 Oct 31;38(11):1968-1979. [doi: 10.1088/1361-6579/aa9047] [Medline: 29087960]
23. Brazendale K, Beets MW, Weaver RG, Perry MW, Tyler EB, Hunt ET, et al. Comparing measures of free-living sleep in school-aged children. Sleep Med 2019 Aug;60:197-201. [doi: 10.1016/j.sleep.2019.04.006] [Medline: 31186212]
24. Brooke SM, An H, Kang S, Noble JM, Berg KE, Lee J. Concurrent validity of wearable activity trackers under free-living conditions. J Strength Cond Res 2017 May;31(4):1097-1106. [doi: 10.1519/JSC.0000000000001571] [Medline: 27465631]
25. Cook JD, Prairie ML, Plante DT. Utility of the Fitbit Flex to evaluate sleep in major depressive disorder: A comparison against polysomnography and wrist-worn actigraphy. J Affect Disord 2017 Aug 01;217:299-305 [FREE Full text] [doi: 10.1016/j.jad.2017.04.030] [Medline: 28448949]
26. Cook JD, Eftekari SC, Dallmann E, Sippy M, Plante DT. Ability of the Fitbit Alta HR to quantify and classify sleep in patients with suspected central disorders of hypersomnolence: A comparison against polysomnography. J Sleep Res 2019 Aug;28(4):e12789. [doi: 10.1111/jsr.12789] [Medline: 30407680]
27. de Zambotti M, Baker FC, Willoughby AR, Godino JG, Wing D, Patrick K, et al. Measures of sleep and cardiac functioning during sleep using a multi-sensory commercially-available wristband in adolescents. Physiol Behav 2016 May 01;158:143-149 [FREE Full text] [doi: 10.1016/j.physbeh.2016.03.006] [Medline: 26969518]
28. de Zambotti M, Goldstone A, Claudatos S, Colrain IM, Baker FC. A validation study of Fitbit Charge 2™ compared with polysomnography in adults. Chronobiol Int 2018 Apr;35(4):465-476. [doi: 10.1080/07420528.2017.1413578] [Medline: 29235907]
29. Dickinson DL, Cazier J, Cech T. A practical validation study of a commercial accelerometer using good and poor sleepers. Health Psychol Open 2016 Jul;3(2):2055102916679012 [FREE Full text] [doi: 10.1177/2055102916679012] [Medline: 28815052]
30. Hakim M, Miller R, Hakim M, Tumin D, Tobias JD, Jatana KR, et al. Comparison of the Fitbit® Charge and polysomnography for measuring sleep quality in children with sleep disordered breathing. Minerva Pediatr 2018 Dec 07:1. [doi: 10.23736/S0026-4946.18.05333-1] [Medline: 30419745]
31. Kang S, Kang JM, Ko K, Park S, Mariani S, Weng J. Validity of a commercial wearable sleep tracker in adult insomnia disorder patients and good sleepers. J Psychosom Res 2017 Jun;97:38-44. [doi: 10.1016/j.jpsychores.2017.03.009] [Medline: 28606497]
32. Kubala AG, Barone Gibbs B, Buysse DJ, Patel SR, Hall MH, Kline CE. Field-based measurement of sleep: Agreement between six commercial activity monitors and a validated accelerometer. Behav Sleep Med 2019 Aug 27:1-16. [doi: 10.1080/15402002.2019.1651316] [Medline: 31455144]
33. Lee H, Lee H, Moon J, Lee T, Kim M, In H, et al. Comparison of wearable activity tracker with actigraphy for sleep evaluation and circadian rest: Activity rhythm measurement in healthy young adults. Psychiatry Investig 2017 Mar;14(2):179-185 [FREE Full text] [doi: 10.4306/pi.2017.14.2.179] [Medline: 28326116]
34. Lee J, Byun W, Keill A, Dinkel D, Seo Y. Comparison of wearable trackers' ability to estimate sleep. Int J Environ Res Public Health 2018 Jun 15;15(6):1265 [FREE Full text] [doi: 10.3390/ijerph15061265] [Medline: 29914050]
35. Liang Z, Chapa Martell MA. Validity of consumer activity wristbands and wearable EEG for measuring overall sleep parameters and sleep structure in free-living conditions. J Healthc Inform Res 2018 Apr 20;2(1-2):152-178. [doi: 10.1007/s41666-018-0013-1]
36. Liu J, Wong WT, Zwetsloot IM, Hsu YC, Tsui KL. Preliminary agreement on tracking sleep between a wrist-worn device Fitbit Alta and consensus sleep diary. Telemed J E Health 2019 Jan 02:1. [doi: 10.1089/tmj.2018.0202] [Medline: 30601109]
37. Mantua J, Gravel N, Spencer RM. Reliability of sleep measures from four personal health monitoring devices compared to research-based actigraphy and polysomnography. Sensors (Basel) 2016 May 05;16(5):646 [FREE Full text] [doi: 10.3390/s16050646] [Medline: 27164110]
38. Maskevich S, Jumabhoy R, Dao PDM, Stout JC, Drummond SPA. Pilot validation of ambulatory activity monitors for sleep measurement in Huntington's disease gene carriers. J Huntingtons Dis 2017;6(3):249-253. [doi: 10.3233/JHD-170251] [Medline: 28968241]
39. Meltzer LJ, Hiruma LS, Avis K, Montgomery-Downs H, Valentin J. Comparison of a commercial accelerometer with polysomnography and actigraphy in children and adolescents. Sleep 2015 Aug 01;38(8):1323-1330 [FREE Full text] [doi: 10.5665/sleep.4918] [Medline: 26118555]
40. Montgomery-Downs HE, Insana SP, Bond JA. Movement toward a novel activity monitoring device. Sleep Breath 2012 Oct;16(3):913-917. [doi: 10.1007/s11325-011-0585-y] [Medline: 21971963]
41. Osterbauer B, Koempel JA, Davidson Ward SL, Fisher LM, Don DM. A comparison study of the Fitbit activity monitor and PSG for assessing sleep patterns and movement in children. J Otolaryngol Adv 2016 Mar 2;1(3):24-35 [FREE Full text] [doi: 10.14302/issn.2379-8572.joa-15-891]
42. Sargent C, Lastella M, Romyn G, Versey N, Miller DJ, Roach GD. How well does a commercially available wearable device measure sleep in young athletes? Chronobiol Int 2018 Jun;35(6):754-758. [doi: 10.1080/07420528.2018.1466800] [Medline: 29750580]

43. Svensson T, Chung U, Tokuno S, Nakamura M, Svensson AK. A validation study of a consumer wearable sleep tracker compared to a portable EEG system in naturalistic conditions. J Psychosom Res 2019 Aug 30;126:109822. [doi: 10.1016/j.jpsychores.2019.109822] [Medline: 31499232]
44. Morgenthaler T, Alessi C, Friedman L, Owens J, Kapur V, Boehlecke B, Standards of Practice Committee, American Academy of Sleep Medicine. Practice parameters for the use of actigraphy in the assessment of sleep and sleep disorders: An update for 2007. Sleep 2007 May;30(4):519-529. [doi: 10.1093/sleep/30.4.519] [Medline: 17520797]
45. Marino M, Li Y, Rueschman MN, Winkelman JW, Ellenbogen JM, Solet JM, et al. Measuring sleep: Accuracy, sensitivity, and specificity of wrist actigraphy compared to polysomnography. Sleep 2013 Dec 01;36(11):1747-1755 [FREE Full text] [doi: 10.5665/sleep.3142] [Medline: 24179309]
46. Fitbit. How do I track my sleep with my Fitbit device? URL: https://help.fitbit.com/articles/en_US/Help_article/1314 [accessed 2018-10-15]
47. Sadeh A. The role and validity of actigraphy in sleep medicine: An update. Sleep Med Rev 2011 Aug;15(4):259-267. [doi: 10.1016/j.smrv.2010.10.001] [Medline: 21237680]
48. Wolfson AR, Carskadon MA, Acebo C, Seifer R, Fallone G, Labyak SE, et al. Evidence for the validity of a sleep habits survey for adolescents. Sleep 2003 Mar 15;26(2):213-216. [doi: 10.1093/sleep/26.2.213] [Medline: 12683482]

Abbreviations

CINAHL: Cumulative Index to Nursing and Allied Health Literature
EBE: epoch-by-epoch
EEG: electroencephalographic
HRV: heart rate variability
MAPE: mean absolute percent error
N1: non-rapid eye movement 1
N2: non-rapid eye movement 2
N3: non-rapid eye movement 3
PLMS: periodic limb movement in sleep
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PSG: polysomnography
REM: rapid eye movement
SE: sleep efficiency
SOL: sleep onset latency
TIB: time in bed
TST: total sleep time
WASO: wake after sleep onset

Edited by G Eysenbach; submitted 16.09.19; peer-reviewed by L Menna-Barreto, A Davoudi, I Mircheva, R Bajpai; comments to author 08.10.19; revised version received 16.10.19; accepted 17.10.19; published 28.11.19.

Please cite as:
Haghayegh S, Khoshnevis S, Smolensky MH, Diller KR, Castriotta RJ
Accuracy of Wristband Fitbit Models in Assessing Sleep: Systematic Review and Meta-Analysis
J Med Internet Res 2019;21(11):e16273
URL: http://www.jmir.org/2019/11/e16273/
doi:10.2196/16273
PMID:31778122

Review

The Effectiveness of Web-Based Interventions Delivered to Children and Young People With Neurodevelopmental Disorders: Systematic Review and Meta-Analysis

Kareem Khan[1], BSc, MSc; Charlotte L Hall[1], BSc, MSc, PhD; E Bethan Davies[1,2], BSc, MSc, PhD; Chris Hollis[1,2,3], MBBS, BSc, DCH, PhD, FRCPsych; Cris Glazebrook[1,2], RGN, PhD, CPsychol

[1]Division of Psychiatry and Applied Psychology, Institute of Mental Health, University of Nottingham, Nottingham, United Kingdom

[2]NIHR MindTech Medtech Co-operative, Institute of Mental Health, University of Nottingham, Nottingham, United Kingdom

[3]NIHR Nottingham Biomedical Research Centre, Nottingham, United Kingdom

Corresponding Author:
Kareem Khan, BSc, MSc
Division of Psychiatry and Applied Psychology
Institute of Mental Health
University of Nottingham
Innovation Park, Triumph Road
Nottingham, NG7 2TU
United Kingdom
Phone: 44 0115 8232438
Email: kareem.khan@nottingham.ac.uk

Abstract

Background: The prevalence of certain neurodevelopmental disorders, specifically autism spectrum disorder (ASD) and attention deficit hyperactivity disorder (ADHD), has been increasing over the last four decades. Nonpharmacological interventions are available that can improve outcomes and reduce associated symptoms such as anxiety, but these are often difficult to access. Children and young people are using the internet and digital technology at higher rates than any other demographic, but although Web-based interventions have the potential to improve health outcomes in those with long-term conditions, no previous reviews have investigated the effectiveness of Web-based interventions delivered to children and young people with neurodevelopmental disorders.

Objective: This study aimed to review the effectiveness of randomized controlled trials (RCTs) of Web-based interventions delivered to children and young people with neurodevelopmental disorders.

Methods: Six databases and one trial register were searched in August and September 2018. RCTs were included if they were published in a peer-reviewed journal. Interventions were included if they (1) aimed to improve the diagnostic symptomology of the targeted neurodevelopmental disorder or associated psychological symptoms as measured by a valid and reliable outcome measure; (2) were delivered on the Web; (3) targeted a youth population (aged ≤18 years or reported a mean age of ≤18 years) with a diagnosis or suspected diagnosis of a neurodevelopmental disorder. Methodological quality was rated using the Joanna Briggs Institute Critical Appraisal Checklist for RCTs.

Results: Of 5140 studies retrieved, 10 fulfilled the inclusion criteria. Half of the interventions were delivered to children and young people with ASDs with the other five targeting ADHD, tic disorder, dyscalculia, and specific learning disorder. In total, 6 of the 10 trials found that a Web-based intervention was effective in improving condition-specific outcomes or reducing comorbid psychological symptoms in children and young people. The 4 trials that failed to find an effect were all delivered by apps. The meta-analysis was conducted on five of the trials and did not show a significant effect, with a high level of heterogeneity detected (n=182 [33.4%, 182/545], 5 RCTs; pooled standardized mean difference=–0.39; 95% CI –0.98 to 0.20; Z=–1.29; *P*=.19 [I^2=72%; *P*=.006]).

Conclusions: Web-based interventions can be effective in reducing symptoms in children and young people with neurodevelopmental disorders; however, caution should be taken when interpreting these findings owing to methodological limitations, the minimal number of papers retrieved, and small samples of included studies. Overall, the number of studies was small and mainly limited to ASD, thus restricting the generalizability of the findings.

Trial Registration: PROSPERO International Prospective Register of Systematic Reviews: CRD42018108824; http://www.crd.york.ac.uk/PROSPERO/display_record.php?ID=CRD42018108824

(J Med Internet Res 2019;21(11):e13478) doi:10.2196/13478

KEYWORDS

online intervention; effectiveness; neurodevelopmental disorders; children and young people; methodology; systematic review

Introduction

Background

Web-based interventions for children and young people (CYP) with physical and psychological problems are relatively new phenomena, with the first trials of internet-delivered therapies being conducted in the late 1990s [1]. However, they are very important developments in the access to health care and treatment for CYP with long-term chronic health conditions. Neurodevelopmental disorders (NDDs) are a group of disorders that typically manifest early in development and are characterized by deficits in cognitive function, motor function, verbal communication, social skills, and behaviors [2]. Common NDDs include autism spectrum disorder (ASD), attention deficit hyperactivity disorder (ADHD), specific learning disorder (including dyscalculia and dyslexia), intellectual disability (ID), and tic disorder ([TD], including tourette syndrome and chronic tic disorder [CTD]) [3]. NDDs frequently co-occur, for example, individuals with ASD often have ID, and many children with ADHD have a specific learning disorder [3]. CYP with NDDs also have complex comorbidities and related symptoms, such as depression and anxiety [4]. There is growing evidence that the impact of NDD is lifelong for many individuals [5], and although exact prevalence rates of NDDs vary considerably between countries, researchers suggest that the prevalence of certain NDDs, specifically ASD and ADHD, has been increasing over the last four decades [6-8].

Psychological therapeutic interventions exist for a range of NDDs. These include therapies to manage NDD symptoms, such as habit reversal therapy for TDs, behavioral therapy to alleviate commonly associated symptoms, such as cognitive behavioral therapy (CBT) for anxiety symptoms, and psychoeducation to facilitate the management of NDDs. Owing to their complexity and chronic nature, pharmacotherapy may often be used as part of a treatment plan [9]. However, pharmacological interventions are considered undesirable for children because of the associated side effects [10]; therefore, psychological treatment is more desirable. A major barrier to psychological treatment is difficulty in accessing appropriately trained therapists, because of the limited numbers of therapists in child mental health services relative to the demand and the uneven geographical distribution of services. It is likely that Web-based therapy can help increase the availability and uptake of evidence-based interventions, offering the opportunity to deliver less therapist-intensive but effective interventions over long distances. Given that Web-based technology is a ubiquitous part of everyday life and young people are by far the highest users [11], Web-delivered therapy is intuitively attractive for CYP.

Web-based interventions are self-guided or therapist-assisted programs with the aim of improving knowledge, providing support, care, or treatment to a diverse population with a range of health problems. In the field of psychological and neurodevelopmental health, Web-based therapeutic interventions have been designed for CYP with a range of problems including ADHD [12], anxiety [13], depression [14], and obsessive-compulsive disorder [15]. These interventions all differ in the type of therapy delivered, their level of participant interaction with the program, number of sessions (dosage), level of trained expert support, structure, modality, and whether there is a parent component or not. However, little is known about what characteristics are integral to efficacious Web-based interventions, especially for CYP. There is some literature in adult populations to suggest that guided Web-based interventions are more efficacious than self-guided or unguided interventions [16], and the most effective interventions tend to be individualized to the user and more intensive [17]. To improve the future developments of Web-based interventions, it would be beneficial to synthesize the evidence for characteristics of effective interventions in CYP to minimize the risk of developing inadequate and ineffective interventions.

A preliminary search conducted in PROSPERO, the Cochrane Database of Systematic Reviews, and the Joanna Briggs Institute (JBI) Database of Systematic Reviews and Implementation Reports indicated that there are no systematic reviews in progress or already published on CYP with NDDs.

Objectives

The objective of this review was to evaluate the effectiveness of Web-based interventions for CYP with NDDs and conduct a meta-analysis of the most effective intervention characteristics (eg, therapist-supported vs stand-alone) with the aim of informing the future development of technologies. The findings will also be useful to health care providers, commissioners, and clinicians in informing future clinical developments in the delivery of care.

Methods

The systematic review was registered on PROSPERO (registration number: CRD42018108824) and conducted in accordance with the JBI methodology for systematic reviews of effectiveness evidence.

Search Strategy

An initial limited scoping search of Medical Literature Analysis and Retrieval System Online (MEDLINE) was undertaken to identify relevant articles. The text words contained in the titles and abstracts of relevant articles and the index and Medical Subject Headings terms describing the articles were used to

develop a full search strategy, which was then tailored for each included information source (see Multimedia Appendix 1 for full search strategy). Search terms were related to NDDs, Web-based interventions, and adolescence.

A total of 6 electronic databases—including PsycINFO, PubMed, EMBASE, Cochrane Central Register of Controlled Trials, Web of Science, and MEDLINE—were searched in August and September 2018. One trial register (ClinicalTrials.gov) was also searched. The reference list of all studies selected for critical appraisal was screened for additional studies, and several specialized journals, publisher websites, and published reviews were hand-searched. As Web-based interventions are a recent development and older interventions will now be obsolete, the year of publication was limited from 2000 to September 5, 2018. There were no restrictions on the language of publication.

Studies were included if they met the following criteria:

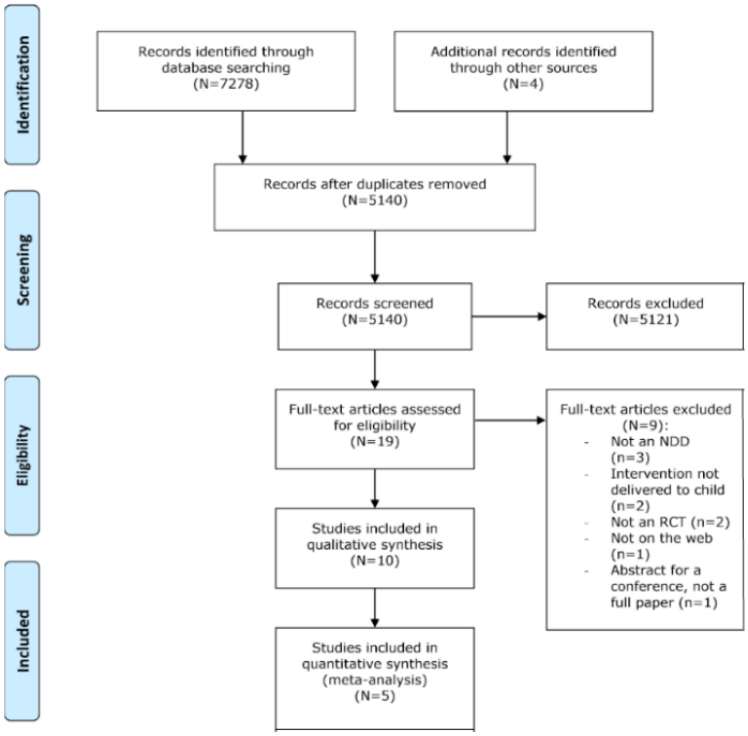
1. The intervention aimed to improve the diagnostic symptomology of the targeted NDD as measured by a valid and reliable outcome measure.
2. The intervention was delivered on the Web via a website, a mobile app, social media, an email, or a personal digital assistant. The intervention could include human support in its delivery.
3. The study was an RCT design and published in a peer-reviewed journal. Trial arms needed to consist of an experimental group compared with no treatment and/or another active intervention or treatment as usual (TAU) or waitlist control.
4. The intervention was targeted at a youth population (aged ≤18 years or reported a mean age of ≤18 years) with a diagnosis or suspected diagnosis of the following NDDs: communication disorders (eg, language disorder and stuttering); ASD; ADHD; specific learning disorder (eg, dyslexia and dyscalculia); motor disorders; TD; other NDDs (eg, NDD associated with prenatal alcohol exposure).

These disorders were selected based on the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) criteria [3].

Secondary outcomes of interest were comorbid or associated psychological symptomology and any adverse events. Papers had to report on either primary or secondary outcomes of interest to be included in this review. Studies were excluded if the intervention was not delivered on the Web or was primarily aimed at the parent or caregiver. Furthermore, we excluded studies where the participants were diagnosed with IDs as intervention characteristics that meet the needs of children with significant IDs would be difficult to generalize to a youth population as a whole. Moreover, studies on NDDs frequently exclude CYP with any form of learning difficulty because of their unique complexity [18].

Once duplicates were removed (n=2142), a total of 5140 titles and abstracts were retrieved. Titles were initially screened against the eligibility criteria by 1 assessor (screening phase, n=4985 ineligible). Subsequently, 155 titles and abstracts were then screened against the eligibility criteria by 2 independent assessors. Any conflicts concerning eligibility were resolved by group discussion. There was agreement on 7 papers to be included, 121 to be excluded, and 27 papers requiring further discussion. Following a discussion between the assessors, the full text of 19 papers was obtained for further analysis and coding. A consensus was reached among the assessors on 9 papers to be excluded, as they did not meet the eligibility criteria, leaving 10 papers for analysis. Figure 1 shows the Preferred Reporting Items for Systematic Reviews and Meta-analyses flowchart [19].

Figure 1. Preferred reporting items for systematic reviews and meta-analyses flowchart outlining the process for systematic review and meta-analysis. NDD: neurodevelopmental disorder; RCT: randomized controlled trial.



Data Extraction

The first assessor extracted the following data from all included studies: specific details about the study (authors, year, number of study arms, location, and Web-based program name), population demographics (sample size, age, and gender), study methods, interventions and comparisons, length of treatment or dosage, condition treated (eg, ASD and ADHD), outcome measures, type of analysis (eg, intention-to-treat [ITT]), and primary and secondary outcomes of significance to the review. These data were extracted and inputted into JBI System for the Unified Management, Assessment and Review of Information (SUMARI) software [20]. Missing data were obtained from the manuscripts, and where these data were not documented, the primary authors were contacted for relevant information.

Assessment of Methodological Quality

A total of 2 independent assessors examined the methodological quality of included studies using the JBI RCT appraisal tool in JBI SUMARI [20]. Further details on the assessment of quality are provided in Multimedia Appendix 2.

Meta-Analysis

Continuous variables were examined using standardized mean differences (SMD) with 95% confidence intervals. Extracted continuous data were tested for normality using skew plots. Random effects meta-analyses were performed to compute overall estimates of treatment outcomes. The effect sizes of the primary studies were presented in a forest plot. Heterogeneity was examined with the I^2 statistic [21]. The I^2 statistic calculates the degree to which there is heterogeneity, with 25% suggesting low heterogeneity, 50% indicating moderate, and 75% indicating the threshold for high heterogeneity. The *Q* statistic was also calculated and provides the statistical significance (*P* value <.05) of heterogeneity.

In the protocol, subgroup analyses were planned to be conducted according to the main intervention characteristics that were shown to be the most effective, for example, therapist support versus no support and parent component versus no parent component. However, because of the low number of included studies in the review, this was deemed unsuitable and is therefore a deviation from the protocol. All data for the meta-analysis were conducted using JBI SUMARI [20].

Results

Study Characteristics

The search generated 10 studies. A total of 5 interventions targeted ASD [22-26], 2 were aimed at CYP with TD [27,28], 1 for ADHD [29], 1 for specific learning disorder (LD) with poor visual-motor integration (VMI) [30], and the other targeting dyscalculia [31]. All but one of the interventions focused on treating the primary diagnosis with the other focusing on treating comorbid anxiety [22]. All studies used the standard RCT design, except for one study, which employed a crossover RCT design [29].

In 5 studies, NDD diagnosis was confirmed by DSM-IV or DSM-5 criteria [23,25-27,29] with the other studies using disorder-specific diagnostic tools [22,24,30,31]. All 10 studies contained 2 trial arms with the intervention being compared with another active intervention, which was not Web-based [27,30,31], TAU, which was either standard therapy or participants were not prevented from using therapy; however they were told not to use any apps designed for ASD therapeutic use [23,26,29] or waitlist control [22,24,25,28]. A summary of the characteristics of each study is shown in Table 1.

Table 1. Characteristics of included studies.

Study	Design, number of arms (N per arm), sample size and study location	Sample demographics and condition treated	Control or comparator group	Outcome measures	Summary of main findings or effect of intervention
Conaughton et al, 2017 [22]	Randomized controlled trial (RCT) 2 arms: Intervention=21, control group=21, N=42, Australia	Children (8-12 years; mean 9.74; 85.7% male) with high-functioning autism spectrum disorder and an anxiety disorder	Waitlist control (WLC)	Anxiety Disorders Interview Schedule: parent and child, Children's Global Assessment Scale, Child Behaviour Checklist, Spence Children's Anxiety Scale–child, satisfaction with treatment	9.5% of the intervention group versus 0% of the WLC group had lost all anxiety diagnoses at postassessment, with 14.3% of the intervention group being free of all anxiety diagnoses at 3-month follow-up; the intervention had a positive effect
Esposito et al, 2017 [23]	RCT 2 arms: Intervention=15, control group=15, N=30, Europe	Children (2-5 years; mean 3.92; 90% male) with Autism Spectrum Disorder (ASD) who followed face-to-face (F2F) applied behavior analysis (ABA) treatment	Treatment as usual (TAU)	Measured attention, imitation of actions with objects, receptive identification of objects	Intervention group, who had daily practice of attention and identification of objects on tablet apps, showed greater progress within standard ABA therapy than the TAU group for all 3 programs investigated; however, this did not exceed the significance level (all *P* values >.05); the intervention had no effect
Fletcher-Watson et al, 2016 [24]	RCT 2 arms: Intervention=27, control group=27, N=54, Europe	Children (<6 years; mean 4.13; 79.6% male) with ASD	WLC	The Autism Diagnostic Observation Schedule, Brief observation of social communication change, MacArthur Communicative Development Inventory (MCDI), Communication and Symbolic Behaviour Scales–Developmental Profile, parent impressions of the app	Change scores on all outcome measures revealed no significant differences between intervention and WLC groups (all *P* values >.05); the intervention had no effect
Fridenson-Hayo et al, 2017 [25]	RCT 2 arms: Intervention=43, control group=40, N=83, Europe	Children (6-9 years; mean 7.29; 79.5% male) with ASD	WLC	Emotion recognition (ER) tasks, Wechsler Intelligence Scale for Children or Wechsler Primary and Preschool Scale of Intelligence, Social Responsiveness Scale, Vineland Adaptive Behaviour Scales (VABS-II)	Pairwise comparisons for the time by group interaction revealed that significant improvement over time was found on all ER tasks for the intervention group but not for the WLC group; the intervention had a positive effect
Whitehouse et al, 2017 [26]	RCT 2 arms: Intervention=41), control group=39, N=80, Australia	Children (<4 years; mean 3.32; 78.7% male) with ASD	TAU	The Autism Treatment Evaluation Checklist (ATEC), The Mullen Scales of Early Learning, VABS-II, MCDI, Communication and Symbolic Behaviour Scales, Repetitive Behaviour Scale-Revised , Behaviour Flexibility Rating Scale	No significant differences were observed between groups for any of the 4 ATEC subscales at either the 3- or 6-month assessments, although the 3-month communication subscale showed a trend toward greater improvement in the intervention group, 2.1 units (95% CI 4.5 to 0.3; *P*=.08); the intervention had no effect

Study	Design, number of arms (N per arm), sample size and study location	Sample demographics and condition treated	Control or comparator group	Outcome measures	Summary of main findings or effect of intervention
Himle et al, 2012 [27]	RCT 2 arms: Intervention=10, comparator group=10, N=20, North America	Children (8-17 years, mean 11.6, 94% male) with tic disorders (TD) or chronic tic disorders (CTD)	F2F Comprehensive Behavioural Intervention for Tics	Yale Global Tic Severity Scale (YGTSS), Clinical Global Impression-Improvement Scale (CGI-I), Parent Tic Questionnaire (PTQ), Treatment Acceptability Questionnaire (TAQ)	The videoconferencing group showed a mean YGTSS reduction of 6.4 points versus 4.2 points for the F2F group at follow-up; both interventions were effective in reducing tics however, there was a slightly better effect on the intervention group at both post-treatment and follow-up compared with the F2F group
Ricketts et al, 2016 [28]	RCT 2 arms: Intervention=12, control group=8, N=20, North America	Children (8-16 years; mean 12.16; 64.9% male) with TD or CTD	WLC	YGTSS, CGI-I, PTQ, Children's Perception of Therapeutic Relationship, Client Satisfaction Questionnaire, TAQ, Videoconferencing Satisfaction Questionnaire	In the intervention group, there was a statistically significant decrease of 7.25 points in YGTSS total scores from baseline to postassessment. In the WLC group, the 1.75-point decrease on the YGTSS total scores from baseline to postassessment was not significant; the intervention had a positive effect
Bul et al, 2016 [29]	Crossover RCT 2 arms: Intervention=88, comparator group=82, N=170, Europe	Children (8-12 years; mean 9.85; 80.6% male) with attention deficit hyperactivity disorder	TAU crossover group	Time management questionnaire, Behaviour Rating Inventory of Executive Function (subscale plan or organize), Social Skills Rating System (subscale cooperation), It's About Time Questionnaire, self-efficacy, satisfaction	Intervention group achieved significantly greater improvements on the primary outcome of time management skills compared with TAU crossover group (parent-reported; P=.004) and on secondary outcomes of responsibility (parent-reported; P=.04), and working memory (parent-reported; P=.02); the intervention had a positive effect
Coutinho et al, 2017 [30]	RCT 2 arms: Intervention=10, comparator group=10, N=20, North America	School-aged children (4-7 years; mean 6.18; 12 males) with a specific learning disorder such as dyspraxia or speech delay with poor visual-motor integration (VMI) skills	Traditional occupational therapy sessions	Beery VMI, Miller function and participation scales, intervention appreciation scale	There were some improvements in VMI skills in both groups; however, the finding was not statistically significant; the intervention had no effect
De Castro et al, 2014 [31]	RCT 2 arms: Intervention=13, control group=13, N=26, South America	Primary school children (7-10 years; mean 8.11; 16 male) with dyscalculia	Traditional teaching techniques	Scholastic Performance Test	The intervention using the virtual environment yielded a significant score improvement (P<.001) with an average score improvement of 5.09 posttest, whereas the control group did not show a statistically significant score improvement (P=.05); the intervention had a positive effect

Modality, Location, and Duration of Intervention

A total of 4 interventions were delivered via apps [23,24,26,30], 2 were serious games [25,29], 2 used videoconferencing [27,28], 1 was a virtual environment with playable games [31], and the other was a Web-based CBT intervention [22]. Most of the interventions were accessed from participants' own homes, except 3 studies where participants were based in a rehabilitation center [30], school [31], and hospital or clinic setting [27]. Interventions either had a varying range of components (ie, tasks to be completed)—2 [24,29], 3 [23], and 4 [25,26] components—or sessions, ranging from 8 [27,28] to 10 [22,30,31] sessions. All trials instructed participants on an optimum length of time to access the intervention: ranging from 5 min per day or 10 min every other day [24], 20 min daily [26] and 30 min per day [23] to approximately 2 hours per week [25], one 60-min session per week [22], 2 40-min sessions per week [30], 60 min twice per week [31], and 65 min 3 times per week [29]. The 2 trials comparing Web-based comprehensive behavioral intervention for tics (CBIT) stated that participants received 6 weekly sessions followed by 2 biweekly sessions [27] and 2 1.5-hour sessions followed by 6 1-hour sessions [28]. The intervention delivery period ranged from 4 [23] to 24 weeks [26], with a median length of 10 weeks.

A summary of the characteristics of each intervention is shown in Table 2.

Table 2. Characteristics of interventions

Study	Intervention, modality, and aim of the intervention	Length or dosage, follow-ups	Therapist supported	Parent component
Conaughton et al [22]	Internet trans diagnostic CBT[a] intervention aimed at improving comorbid anxiety symptoms	10 weeks, 10 sessions—one 60-min session per week	Yes	Yes
Esposito et al [23]	Tablet apps aimed at improving attention and identification of objects	4 weeks, 3 app components—30 min daily	Yes	Yes
Fletcher-Watson [24]	iPad app aimed to improve social communication skills	2-months, 2 parts–5 min per day, or 10 min every other day	No	No
Fridenson-Hayo et al [25]	An internet-based serious game aimed at improving emotion recognition	8-12 weeks, 4 components—2 hours per week	No	Yes
Whitehouse et al [26]	iPad app aimed at improving developmental skills relevant to autism	6 months, 4 components–20 min per day	No	Yes
Himle et al [27]	Internet-accessed videoconference aimed at improving tic severity	10 weeks—6 weekly sessions followed by 2 biweekly sessions	Yes	Yes
Ricketts et al [28]	Internet-accessed videoconference (Skype) aimed at improving tic severity	10 weeks—2 1.5-hour sessions followed by 6 1-hour sessions	Yes	Yes
Bul et al [29]	An internet-based serious game aimed at improving time management and planning skills	10 weeks, 2 game components—65 min approximately 3 times per week	No	No
Coutinho et al [30]	Multiple iPad apps aimed at improving visual motor skills	10 weeks, minimum of 8 and maximum of 12 sessions—2 40-min sessions per week	No	No
De Castro et al [31]	Internet-accessed virtual environment aimed at improving mathematical skills	5 weeks, 10 sessions—60 min twice a week	No	No

[a]CBT: cognitive behavioral therapy.

Use of Human and Technical Support

In total, 4 interventions were therapist assisted [22,23,27,28]; however, all these differed in the level of involvement of the therapist within the interventions. The contacts ranged from once weekly contact [22] 2 hours per week [23], and the 2 trials of CBIT were exclusively therapist-delivered [27,28].

One of the major factors that developers need to consider when creating Web-based intervention is the ease with which nontechnologically advanced individuals can access and use the program. Thus, it is crucial to provide technical support as and when needed. In total, 7 of the 10 included studies reported the use of technical support. In 2 trials [22,28], participants had weekly access to a therapist who was able to offer any technical assistance within the sessions. One trial [30] took place within a rehabilitation center with an occupational therapist (OT) constantly present to offer any assistance. Two trials reported the use of monitoring phone calls from research personnel to check for any issues, which were offered either fortnightly [26] or once a week [25]. In both of these trials, parents were also encouraged to contact research staff with any queries or issues in between monitoring calls. In one trial [27], research personnel were available to manage any technical difficulties. In the other trial [23], parents were fully trained in the apps by research staff and were taught how to handle technical difficulties.

Participant Characteristics

A total of 545 participants consented and were randomized to a trial arm. Sample sizes ranged from 20 [27,28,30] to 170 [29] participants. A total of 4 trials had sample sizes of >50 participants [24-26,29]. Overall, 523 participants were explicitly included in analyses. A total of 5 studies stated that the analysis was conducted on participants who completed pre- and postintervention measures only [23,25,27,30,31], whereas 5 conducted ITT analyses [22,24,26,28,29]. All 10 trials reported participant dropout or withdrawal data, with dropout rates ranging from 0% [23,28,31] to 18% (n=31) of the sample [29]. Reasons for participant withdrawal included lack of motivation or disinterest [25,29], lack of enjoyment with the intervention [24,26], and personal reasons [26].

In the 10 trials, participants ranged in age from 2 to 17 years, with a mean age ranging from 3.32 to 12.16 years. Males were the majority in all studies, with gender balance varying from 62.5% [31] to 94% [27] of the sample being male. A total of 4 trials were conducted in Europe [23-25,29], 3 in North America [27,28,30], 2 in Australia [22,26], and 1 in South America [31].

Provider Characteristics

Most of the trials recruited participants from clinics [22-25,29], with 3 studies [25,26,28] recruiting via advertisements and 1 study [28] recruiting participants through solicitations mailed to health care professionals. One study [26] recruited participants

through referrals from diagnosing clinicians, and another study [22] utilized referrals through general practitioners, mental health professionals, school guidance officers, teachers, parents, and media publicity.

Adverse Events and Outcome Measures

Only 1 study [29] explicitly stated that they recorded and reported adverse events. The crossover trial investigating the effects of a serious game as an adjunct to TAU for children with ADHD reported 10 adverse events in the trial that could be related to the intervention, and parents, teachers, or participants themselves reported these. Adverse events were registered as mild (n=5) or moderate (n=5) in severity and examples included pain in the fingers, irritability, and headache. One participant could not concentrate at school and therefore discontinued from the trial because of this adverse event; however, no serious adverse events were reported.

It was estimated that the outcome measurement battery ranged from 16 [24] to 175 items [26] at each time point of the studies. The estimated median number of questions administered to participants was 56 items (see Multimedia Appendix 2 for more details).

Methodological Quality and Risk of Bias

The JBI Critical Appraisal Checklist for RCTs provided a framework for scoring the quality of the included studies by addressing different aspects of the research such as randomization, allocation concealment, blinding, and follow-up data. The methodological quality of included studies was felt to be moderate, mostly because of trials providing insufficient details or being unclear in their reporting (see Table 3). Only 5 of 10 studies reported their randomization methodology [22,24,28-30]. Blinding was the main issue of quality in included studies. A total of 6 trials stated that participants were not blind to treatment assignment with the other 4 trials being unclear in their reporting. Only 1 study [23] reported that those delivering treatments were blind to treatment assignment with the others stating researchers delivering treatment were either not blinded or it was unclear. Half of the trials [22-24,26,27] reported outcome assessors were blind to treatment assignment with all of these studies employing independent researchers to carry out assessments.

Table 3. Critical appraisal of included studies.

Study	Q1[a]	Q2[b]	Q3[c]	Q4[d]	Q5[e]	Q6[f]	Q7[g]	Q8[h]	Q9[i]	Q10[j]	Q11[k]	Q12[l]	Q13[m]
Conaughton et al [22]	Yes	Yes	Yes	Unclear	No	Yes	No	Yes	Yes	No	Yes	Yes	Yes
Esposito et al [23]	Unclear	Unclear	Yes	Unclear	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fletcher-Watson et al [24]	Yes	Yes	Yes	No	Unclear	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fridenson-Hayo et al [25]	Unclear	Unclear	Yes	No	No	Unclear	Yes	Yes	Unclear	Yes	Yes	Yes	Yes
White-house et al [26]	Unclear	Unclear	Yes	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Himle et al [27]	Unclear	Unclear	Yes	No	No	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Ricketts et al [28]	Yes	Unclear	Yes	No	No	Unclear	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bul et al [29]	Yes	No	Yes	No	Unclear	Unclear	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Coutinho et al [30]	Yes	Unclear	Yes	Unclear	Unclear	Unclear	Yes	Yes	Yes	Yes	Yes	Yes	Yes
De Castro et al [31]	Unclear	Unclear	Yes	Unclear	Unclear	Unclear	Yes	Yes	Unclear	Yes	Yes	Yes	Yes
Number that met the criteria (%)	50	20	100	0	10	50	90	100	70	90	100	100	100

[a]Q1: True randomization.

[b]Q2: Allocation concealed.

[c]Q3: Treatment groups similar at the baseline.

[d]Q4: Participants blind to treatment.

[e]Q5: Those delivering intervention blind to treatment.

[f]Q6: Outcome assessors blind to treatment.

[g]Q7: Treatment groups treated identically.

[h]Q8: Follow-up complete and if not, differences between groups adequately described and analyzed.

[i]Q9: Participants analyzed in the groups to which they were randomized.

[j]Q10: Outcomes measured in the same way for groups.

[k]Q11: Outcomes measured reliably.

[l]Q12: Appropriate statistical analysis.

[m]Q13: Appropriateness of trial design and any deviations from RCT design accounted for.

Effectiveness of Web-Based Interventions

Of 10 trials, 6 trials found that Web-based interventions were effective in reducing NDDs or associated symptoms in CYP [22,25,27-29,31]; 2 were serious games, 2 were delivered by videoconferencing, 1 was a virtual environment, and the other was an internet-delivered CBT intervention. Targeted NDD conditions of the effective interventions included ASD [22,25], TD [27,28], ADHD [29], and dyscalculia [31]. All but 2 of the effective interventions were delivered over a period of 10 weeks, and these 2 were delivered over 5 weeks with 10 sessions [31] and 8 to 12 weeks with 4 components [25]. The 4 trials, which did not find that Web-based interventions had an effect on NDD symptoms, were all delivered by apps [23,24,26,30]. All but one of these was designed for CYP with ASD, the other being designed for specific LD with VMI [30].

Primary Outcomes

Of 10 interventions, 4 interventions in the included studies were aimed at a youth population with ASD; however, just one [25] of these trials found that Web-based interventions were effective. In the study by Fridenson-Hayo et al [25], children with ASD who received an internet-based serious game improved in ER tasks compared with the WLC group who received TAU. A total of 3 studies [23,24,26] comparing iPad or tablet apps with

WLC/TAU groups for children with ASD found no difference in outcome between the groups.

Both studies evaluating the effectiveness of internet-delivered CBIT via videoconferencing for young people with TD/CTD showed it could be effective for reducing tic symptomology. Overall, the studies were of similar design but used different comparators with Himle et al [27] using F2F CBIT in their study whereas WLC was utilized in a study by Ricketts et al [28]. The YGTSS was the main primary measure in both trials.

There were 3 other studies that looked to improve primary symptoms in CYP, and these were targeted at CYP with NDDs other than ASD or TD. One study showed improvements in time management skills for children with ADHD [29], and another study found improvements in mathematical skills for children with dyscalculia [31]. The other study found no effect in VMI scores [30]. Secondary outcomes are discussed in Multimedia Appendix 2.

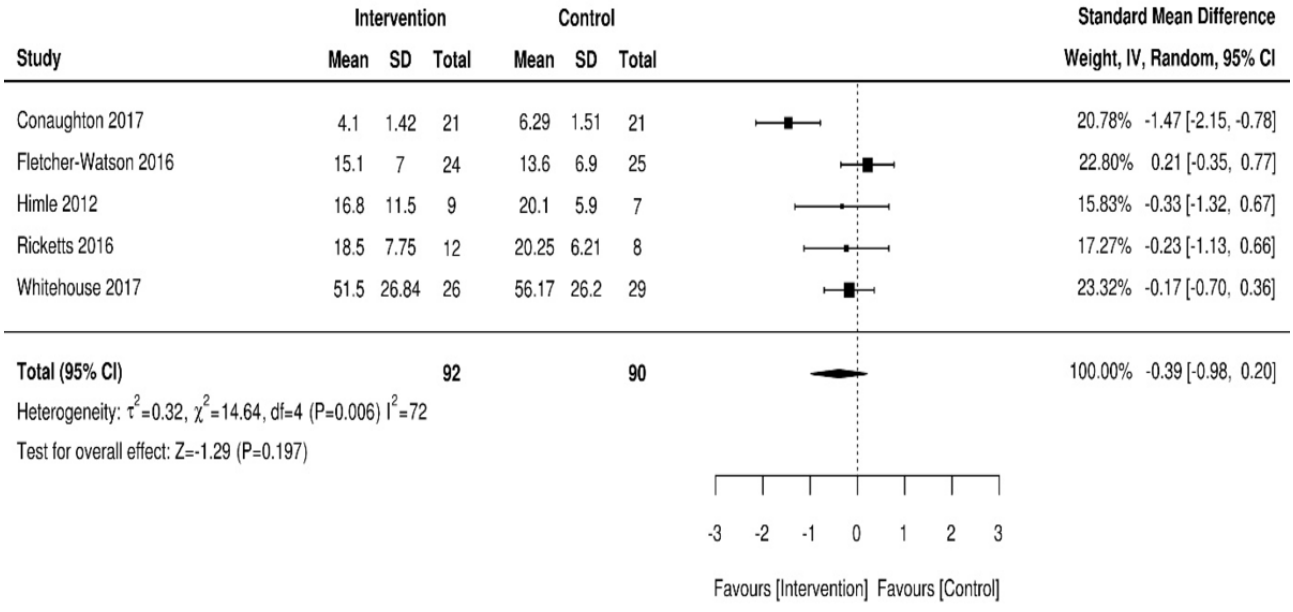
Satisfaction or Acceptability of the Intervention

A total of 4 trials included participant satisfaction measures [22,24,26,29] and 2 trials administered participant acceptability questionnaires [27,28]. In the study by Bul et al [29], both children and parents reported moderate to high satisfaction with receiving the serious game intervention. In the study by Conaughton et al [22], children and parents reported moderate levels of satisfaction following treatment. In the study by Fletcher-Watson et al [24], parents gave verbal comments on the app and what they perceived to be their child's response to it. Replies were categorized as *Positive*, *Mixed*, or *Negative*, and there were positive responses to questions on overall experience with the app, whether the child and parent liked the app, and ease of use. In the other study to measure participant satisfaction [26], caregivers of children in the Therapy Outcomes By You (TOBY) intervention group were asked to list up to 3 features that they liked or disliked about the app. The most frequent *like* statement related to TOBY providing a helpful therapy-planning tool. Other common statements were that TOBY was easy to use and that the app provided a positive learning experience for their child with an attractive structure and layout. The most common *dislike* statement was that the offline iPad activities were too time-consuming to prepare. The 2 trials evaluating VC administered CBIT [27,28] gathered acceptability ratings from participants. In both studies, children and parents gave high acceptability ratings for the intervention.

Meta-Analysis

In studies that used a valid and reliable outcome measurement of NDD and associated symptoms, a meta-analysis was undertaken. All outcomes were continuous and scale-based and were extracted as endpoint average scores with lower scores indicating less severe symptomology. The outcomes combined for the meta-analysis were anxiety [22], social communication [24], developmental skills [26], and tic severity [27,28]. Negative SMD values support the intervention in the presented analyses. Figure 2 shows the forest plot for the data.

Figure 2. Forest plot of postintervention neurodevelopmental disorder outcomes for intervention compared with controls.



A total of 5 trials investigated the effects of Web-based interventions on NDD symptoms using a valid, standardized outcome measure to explore symptom reduction. Within the 5 trials, neither intervention nor control was favored, with a high level of heterogeneity detected: 182/545 (33.4%), 5 RCTs, pooled SMD=–0.39; 95% CI –0.98 to 0.20; Z=–1.29; *P*=.19 (I^2=72%; *P*=.006).

Discussion

Principal Findings

We set out to evaluate whether RCT evidence showed Web-based interventions were effective for CYP with NDDs and/or associated symptoms. Our review retrieved 10 studies in total. A further meta-analysis was conducted on 5 of the 10 studies. Most of the interventions targeted ASD in CYP. Overall,

the meta-analysis indicated no difference between the intervention and control groups; however, with 6 of the 10 retrieved papers showing a positive effect, the findings suggest that Web-based interventions can be effective in reducing NDD symptoms in CYP. However, the evidence is inconclusive owing to the limited number of retrieved studies and small sample sizes in included trials. The findings indicate the need for further research in the use of Web-based interventions aimed at CYP with NDDs.

Furthermore, one of our initial aims was to evaluate the main characteristics of effective Web-based interventions. A parent component as an adjunct to the main intervention was utilized in 4 of the 6 effective trials, indicating the potential importance of assisted interventions and in line with previous research [16,32,33]. Having a parent component within the interventions is unsurprising given the young age of participants in the included studies. It is more likely that younger children will require some form of parental assistance with digitized interventions and, more generally, therapeutic interventions. Indeed, Thirwall et al [34] found that younger children showed a greater improvement in anxiety symptoms having received a parent-delivered CBT intervention. From this review, it is unclear whether a therapist-supported Web-based intervention is more efficacious than one without, as only half of the effective interventions were therapist supported. Another important characteristic to consider is the length of the intervention. A total of 5 of the 6 effective interventions were delivered over a period of 10 to 12 weeks, with the other having 10 sessions delivered over 5 weeks. This suggests that 10 to 12 weeks/sessions is the optimum length for a Web-based intervention. However, given the high heterogeneity between the Web-based interventions and number of multifaceted aspects to these interventions in this review, caution should be taken when trying to establish certain characteristics that may be relevant in determining effectiveness.

All 4 of the included interventions delivered by apps were unsuccessful in yielding statistically significant outcomes. This suggests apps may not be a promising platform for delivering therapeutic interventions, at least to CYP with NDDs. Indeed, recent systematic reviews [35,36], have shown there is inconclusive evidence on the efficacy of mobile apps utilized as health interventions, despite the high user acceptability ratings of smartphone apps. One interpretation of this finding is that because apps are a new phenomenon—the first mobile apps being developed in 2008 with the advent of Apple's App Store [37]—little is known about their mechanisms of impact, especially in the health care domain. There are over 10,000 mental health apps commercially available [38], with 52% of smartphone owners using their phones for health purposes and 19% using health apps [39], it is clear that more high-quality research needs to be conducted. As 3 of the 4 apps that found no effect were targeted at CYP with ASD, another interpretation of this finding could be that apps are an insufficient modality for producing positive outcomes in autism-related disorders. This corroborates the results of a study conducted by Grynszpan et al [40]. They found that adolescents with ASD performed poorly on rich multimedia interfaces, such as apps, as they lacked the required initiative in organizing information given within the multimodal sources.

Half of the included interventions were delivered to CYP with ASD, and much of the research to date evaluating digital technologies administered to NDDs has focused on ASD [41-43]. A possible explanation for this is that computer technology can help compensate verbal and social interaction difficulties and enable facilitation of exchanges between people with ASD, experts, and others [44]. The vast potential of technology for ASD has been realized by researchers, as technologies can enable new ways of communicating for people with ASD, socializing, and even learning. Despite this, many studies still lack scientific rigor to allow for concrete support for the use of technology in aiding people with ASD [42]. In this review, 2 of the 5 RCTs found that Web-based interventions were effective for CYP with ASD and one of these targeted CYP with HFASD who had a comorbid diagnosis of an anxiety disorder.

The RCTs included in this review were assessed as being of acceptable quality for a review of effectiveness. However, the main methodological issues centered on the lack of blinding of participants and of those delivering treatment. All studies had a control group, which was either active or inactive, with half of the trials using valid, standardized outcome measures. Most trials had low attrition rates thus improving the overall quality of the included studies. Only 1 of the 10 trials explicitly recorded and reported adverse events [29]. They reported on 10 adverse events that could be related to the intervention however, none were regarded as serious. Insufficient reporting of adverse events in psychological treatments has been documented in the literature [45], and it is clear that future trials should be more explicit in their reporting.

Limitations

Some limitations of the review and meta-analysis need to be considered. A major limitation is the minimal number of studies retrieved meaning that any conclusions drawn from this review must be met with caution. To provide an expansive overview of the effectiveness of Web-based interventions for CYP, we included trials targeting a myriad of NDDs, which may have equilibrated disorder-specific effects of Web-based interventions. As there were very few RCTs evaluating the effectiveness of Web-based interventions in CYP with NDDs, it would have been impractical to carry out a review focusing on 1 NDD only. We could have increased the number of NDDs by also including trials focusing on CYP with learning disabilities; however, this would have further increased the heterogeneity and added to the problems of generalizability owing to the complexity of this particular population. The search was conducted on multiple databases and updated through a repeated search, thus ensuring a comprehensive overview of the topic. A particular strength of this review is that we had 2 independent reviewers screening relevant papers, with discrepancies between the reviewers discussed. This ensured a structured, meticulous approach was undertaken in study selection, therefore, improving review quality.

For the meta-analysis, we could only include data from 5 of the 10 trials, meaning the pool of data from included interventions

was small and limited the overall power. Moreover, there was a high level of heterogeneity detected in the meta-analysis, which may have been because of the types of comparison with the interventions or differences in baseline symptomology [46]. There is mixed literature on whether a meta-analysis should be conducted at all in the event of high heterogeneity; however, experts recommend using the random effects model [21,47] that was used in this review. Finally, a major strength of this review is that it is based on *a priori* protocol that decreases the potential for reviewer bias.

When interpreting the findings, some inherent methodological issues of the included studies must also be considered, as methodological flaws of the primary trials can have a considerable impact on the review results. One intrinsic methodological limitation of many therapeutic intervention trials is the lack of blinding of participants and those delivering treatment [48], thus introducing a high risk of bias. As already mentioned, most of the included trials had very small sample sizes, which makes the generalization of findings highly problematic. All interventions used different content and modalities of delivery, which could have affected participant interaction and consequently, effectiveness [49]. Another limitation is with the RCT design itself. Given that the most effective interventions are individualized to the user [17], this is often difficult to assess using an RCT design, meaning the interventions reviewed mostly fell short on this dimension.

Gender balance was a potential issue of bias in included studies, as most of the trials had more male participants than female. However, this is not surprising given that NDDs are more common in males than females [50]. Baseline symptomology was also a potential source of bias, as this may have caused difficulties comparing intervention effectiveness in improving NDD outcomes. Some trials recruited participants with minimal symptoms, whereas others recruited those experiencing high levels of NDD symptoms. Despite these limitations, the overall reporting of the included trials was of a high standard and methodologically sound.

Implications for Practice

As some of the interventions found positive outcomes, health care professionals working with CYP may want to consider utilizing Web-based and digital resources to support their patients, especially those with tics. The National Health Service (NHS) has already developed improving access to psychological therapy services for young people with mental health problems and is aiming to incorporate this into practice nationwide within the coming years [51]. If this is successful in reducing the burden on health care services and is shown to be cost-effective, this could lead to promising new developments for digital resources to be used on other populations. None of the included studies assessed the cost-effectiveness of Web-based interventions, which is likely to be an important consideration for policymakers. All the efficacious interventions in this review contained an element of human interaction, either with a real person by videoconferencing or a simulated person in a virtual environment or serious game. The best improvement in outcomes, therefore, may be achieved through a combination of Web-based interventions and human support. As technology evolves rapidly, future Web-based interventions will be more dynamic, perhaps including real-time clinician or therapist input and integrated synchronous crisis support. A promising new development is the use of virtual reality, which has had positive results on children with ADHD [52], adults with anxiety disorders [53], and a range of other mental health problems [54]. Developers could utilize virtual reality to its full effect and enable a simulated, life-like human therapist to support CYP with NDDs and common comorbidities, thus cutting waiting lists while improving outcomes.

Implications for Research

Future studies of Web-based interventions for CYP with NDDs must have larger sample sizes to generate a reasonable degree of statistical power and allow for an increase in generalizability. They must also consider including long-term follow-up assessments to evaluate whether effects are maintained over a prolonged period. A cost-effectiveness evaluation would also be appropriate and much needed in future research. Furthermore, qualitative feedback in the form of a process evaluation would be useful in addressing the intervention's mechanisms of impact and usability.

Our review found multiple methodological issues with the included trials. Sources of high risk of bias in the RCTs included failure to blind participants and personnel to the Web-based intervention and inadequate reporting of allocation concealment. Failing to blind participants, which can be difficult in Web-based intervention studies, can lead to the *digital placebo effect* [55]. One possible way of mediating this effect in future studies is to create a sham or static Web-based program for control groups, therefore, reducing the risk of the digital placebo effect. As mentioned, individualized interventions are often the most effective; however, RCT designs are inadequate in assessing the individualized dimension of interventions, therefore future studies should focus on conducting single case experimental designs to measure this [56,57].

Conclusions

Technological advances and mobile device popularity have huge potential to improve outcomes in CYP with NDDs and comorbid psychological problems. Overall, this study suggests that Web-based interventions can be beneficial in improving symptoms in this population; however, because of the small number of RCTs yielded and several methodological limitations in the included studies, mean findings must be considered with caution. There need to be more studies with larger sample sizes assessing the effectiveness of Web-based interventions for CYP. Furthermore, a qualitative evaluation of the intervention is encouraged in future work to provide bespoke Web-based interventions for youth populations.

Acknowledgments

The PhD is funded by National Institute for Health Research (NIHR) MindTech MedTech Cooperative and NIHR Nottingham BRC Mental Health & Technology Theme. The PhD explores the process evaluation of the Online Remote Behavioral Intervention for Tics Trial, which is funded by the NIHR Health Technology Assessment (Ref 16/19/02).

The PhD studentship is funded by the NIHR MindTech MedTech Cooperative from whom EBD, CH, and CG received financial support. This study is funded by the National Institute for Health Research. The views represented are the views of the authors alone and do not necessarily represent the views of the Department of Health in England, NHS, or the National Institute for Health Research.

CLH acknowledges receiving financial support from NIHR Collaborations for Leadership in Applied Health Research and Care East Midlands.

Authors' Contributions

KK conducted the literature searches and extracted, tabulated, and interpreted data, contributed to the conception and design of the review, and wrote the manuscript. CLH and EBD aided in the protocol development, performed the second review of the papers, and critically reviewed and revised the manuscript. CG contributed to the conception and design of the review, critically reviewed, and revised the manuscript. CH approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Full search strategy.
[PDF File (Adobe PDF File), 135 KB - jmir_v21i11e13478_app1.pdf]

Multimedia Appendix 2
Further information.
[PDF File (Adobe PDF File), 59 KB - jmir_v21i11e13478_app2.pdf]

References

1. Barak A. Psychological applications on the internet: a discipline on the threshold of a new millennium. Appl Prev Psychol 1999 Sep;8(4):231-245. [doi: 10.1016/S0962-1849(05)80038-1]
2. Ahn S, Hwang S. Cognitive rehabilitation with neurodevelopmental disorder: a systematic review. NeuroRehabilitation 2017;41(4):707-719. [doi: 10.3233/NRE-172146] [Medline: 28946583]
3. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. Fifth Edition. Washington, DC: American Psychiatric Association Publishing; 2013.
4. King BH. Psychiatric comorbidities in neurodevelopmental disorders. Curr Opin Neurol 2016 Apr;29(2):113-117. [doi: 10.1097/WCO.0000000000000299] [Medline: 26840045]
5. Cleaton MA, Kirby A. Why do we find it so hard to calculate the burden of neurodevelopmental disorders. J Child Dev Disord 2018 Mar;4(3):330-338. [doi: 10.4172/2472-1786.100073]
6. Grandjean P, Landrigan PJ. Developmental neurotoxicity of industrial chemicals. Lancet 2006 Dec 16;368(9553):2167-2178. [doi: 10.1016/S0140-6736(06)69665-7] [Medline: 17174709]
7. Rutter M. Incidence of autism spectrum disorders: changes over time and their meaning. Acta Paediatr 2005 Jan;94(1):2-15. [doi: 10.1111/j.1651-2227.2005.tb01779.x] [Medline: 15858952]
8. McCarthy S, Wilton L, Murray ML, Hodgkins P, Asherson P, Wong IC. The epidemiology of pharmacologically treated attention deficit hyperactivity disorder (ADHD) in children, adolescents and adults in UK primary care. BMC Pediatr 2012 Jun 19;12:78 [FREE Full text] [doi: 10.1186/1471-2431-12-78] [Medline: 22712630]
9. Homberg JR, Kyzar EJ, Scattoni ML, Norton WH, Pittman J, Gaikwad S, et al. Genetic and environmental modulation of neurodevelopmental disorders: translational insights from labs to beds. Brain Res Bull 2016 Jul;125:79-91. [doi: 10.1016/j.brainresbull.2016.04.015] [Medline: 27113433]
10. Whittington C, Pennant M, Kendall T, Glazebrook C, Trayner P, Groom M, et al. Practitioner review: treatments for Tourette syndrome in children and young people - a systematic review. J Child Psychol Psychiatry 2016 Sep;57(9):988-1004. [doi: 10.1111/jcpp.12556] [Medline: 27132945]
11. Anderson M, Jiang J. Pew Research Center. 2018. Teens, Social Media & Technology 2018 URL: http://www.pewinternet.org/2018/05/31/teens-social-media-technology-2018/ [accessed 2019-01-10] [WebCite Cache ID 75K43Vi9Q]

12. Dovis S, van der Oord S, Wiers RW, Prins PJM. Improving executive functioning in children with ADHD: training multiple executive functions within the context of a computer game. a randomized double-blind placebo controlled trial. PLoS One 2015;10(4):e0121651 [FREE Full text] [doi: 10.1371/journal.pone.0121651] [Medline: 25844638]
13. Spence S, Donovan C, March S, Gamble A, Anderson R, Prosser S, et al. A randomized controlled trial of online versus clinic-based CBT for adolescent anxiety. J Consult Clin Psychol 2011 Oct;79(5):629-642. [doi: 10.1037/a0024512] [Medline: 21744945]
14. Whittaker R, Merry S, Stasiak K, McDowell H, Doherty I, Shepherd M, et al. MEMO--a mobile phone depression prevention intervention for adolescents: development process and postprogram findings on acceptability from a randomized controlled trial. J Med Internet Res 2012 Jan 24;14(1):e13 [FREE Full text] [doi: 10.2196/jmir.1857] [Medline: 22278284]
15. Lenhard F, Andersson E, Mataix-Cols D, Rück C, Vigerland S, Högström J, et al. Therapist-guided, internet-delivered cognitive-behavioral therapy for adolescents with obsessive-compulsive disorder: a randomized controlled trial. J Am Acad Child Adolesc Psychiatry 2017 Jan;56(1):10-9.e2 [FREE Full text] [doi: 10.1016/j.jaac.2016.09.515] [Medline: 27993223]
16. Baumeister H, Reichler L, Munzinger M, Lin J. The impact of guidance on internet-based mental health interventions — a systematic review. Internet Interv 2014 Oct;1(4):205-215. [doi: 10.1016/j.invent.2014.08.003]
17. Rogers MA, Lemmen K, Kramer R, Mann J, Chopra V. Internet-delivered health interventions that work: systematic review of meta-analyses and evaluation of website availability. J Med Internet Res 2017 Mar 24;19(3):e90 [FREE Full text] [doi: 10.2196/jmir.7111] [Medline: 28341617]
18. Bishop DV. Which neurodevelopmental disorders get researched and why? PLoS One 2010 Nov 30;5(11):e15112 [FREE Full text] [doi: 10.1371/journal.pone.0015112] [Medline: 21152085]
19. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
20. Munn Z. Software to support the systematic review process: the Joanna Briggs Institute System for the Unified Management, Assessment and Review of Information (JBI-SUMARI). JBI Database System Rev Implement Rep 2016 Oct;14(10):1. [doi: 10.11124/JBISRIR-2016-002421] [Medline: 27846109]
21. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med 2002 Jun 15;21(11):1539-1558. [doi: 10.1002/sim.1186] [Medline: 12111919]
22. Conaughton RJ, Donovan CL, March S. Efficacy of an internet-based CBT program for children with comorbid high functioning autism spectrum disorder and anxiety: a randomised controlled trial. J Affect Disord 2017 Aug 15;218:260-268. [doi: 10.1016/j.jad.2017.04.032] [Medline: 28477505]
23. Esposito M, Sloan J, Tancredi A, Gerardi G, Postiglione P, Fotia F, et al. Using tablet applications for children with autism to increase their cognitive and social skills. J Spec Educ Technol 2017 Jul 18;32(4):199-209. [doi: 10.1177/0162643417719751]
24. Fletcher-Watson S, Petrou A, Scott-Barrett J, Dicks P, Graham C, O'Hare A, et al. A trial of an iPad™ intervention targeting social communication skills in children with autism. Autism 2016 Oct;20(7):771-782 [FREE Full text] [doi: 10.1177/1362361315605624] [Medline: 26503990]
25. Fridenson-Hayo S, Berggren S, Lassalle A, Tal S, Pigat D, Meir-Goren N, et al. 'Emotiplay': a serious game for learning about emotions in children with autism: results of a cross-cultural evaluation. Eur Child Adolesc Psychiatry 2017 Aug;26(8):979-992. [doi: 10.1007/s00787-017-0968-0] [Medline: 28275895]
26. Whitehouse AJ, Granich J, Alvares G, Busacca M, Cooper MN, Dass A, et al. A randomised controlled trial of an iPad-based application to complement early behavioural intervention in autism spectrum disorder. J Child Psychol Psychiatry 2017 Sep;58(9):1042-1052. [doi: 10.1111/jcpp.12752] [Medline: 28543302]
27. Himle MB, Freitag M, Walther M, Franklin SA, Ely L, Woods DW. A randomized pilot trial comparing videoconference versus face-to-face delivery of behavior therapy for childhood tic disorders. Behav Res Ther 2012 Sep;50(9):565-570. [doi: 10.1016/j.brat.2012.05.009] [Medline: 22743661]
28. Ricketts EJ, Goetz AR, Capriotti MR, Bauer CC, Brei NG, Himle MB, et al. A randomized waitlist-controlled pilot trial of voice over internet protocol-delivered behavior therapy for youth with chronic tic disorders. J Telemed Telecare 2016 Apr;22(3):153-162 [FREE Full text] [doi: 10.1177/1357633X15593192] [Medline: 26169350]
29. Bul KC, Kato PM, van der Oord S, Danckaerts M, Vreeke LJ, Willems A, et al. Behavioral outcome effects of serious gaming as an adjunct to treatment for children with attention-deficit/hyperactivity disorder: a randomized controlled trial. J Med Internet Res 2016 Feb 16;18(2):e26 [FREE Full text] [doi: 10.2196/jmir.5173] [Medline: 26883052]
30. Coutinho F, Bosisio M, Brown E, Rishikof S, Skaf E, Zhang X, et al. Effectiveness of iPad apps on visual-motor skills among children with special needs between 4y0m-7y11m. Disabil Rehabil Assist Technol 2017 May;12(4):402-410. [doi: 10.1080/17483107.2016.1185648] [Medline: 27377131]
31. de Castro MV, Bissaco MA, Panccioni BM, Rodrigues SC, Domingues AM. Effect of a virtual environment on the development of mathematical skills in children with dyscalculia. PLoS One 2014;9(7):e103354 [FREE Full text] [doi: 10.1371/journal.pone.0103354] [Medline: 25068511]
32. Johansson R, Andersson G. Internet-based psychological treatments for depression. Expert Rev Neurother 2012 Jul;12(7):861-9; quiz 870. [doi: 10.1586/ern.12.63] [Medline: 22853793]

33. Richards D, Richardson T. Computer-based psychological treatments for depression: a systematic review and meta-analysis. Clin Psychol Rev 2012 Jun;32(4):329-342. [doi: 10.1016/j.cpr.2012.02.004] [Medline: 22466510]
34. Thirlwall K, Cooper P, Creswell C. Guided parent-delivered cognitive behavioral therapy for childhood anxiety: predictors of treatment response. J Anxiety Disord 2017 Jan;45:43-48 [FREE Full text] [doi: 10.1016/j.janxdis.2016.11.003] [Medline: 27930939]
35. Byambasuren O, Sanders S, Beller E, Glasziou P. Prescribable mHealth apps identified from an overview of systematic reviews. NPJ Digit Med 2018;1:12 [FREE Full text] [doi: 10.1038/s41746-018-0021-9] [Medline: 31304297]
36. Payne HE, Lister C, West JH, Bernhardt JM. Behavioral functionality of mobile apps in health interventions: a systematic review of the literature. JMIR Mhealth Uhealth 2015 Feb 26;3(1):e20 [FREE Full text] [doi: 10.2196/mhealth.3335] [Medline: 25803705]
37. Yoo JH. The meaning of information technology (IT) mobile devices to me, the infectious disease physician. Infect Chemother 2013 Jun;45(2):244-251 [FREE Full text] [doi: 10.3947/ic.2013.45.2.244] [Medline: 24265976]
38. Torous J, Andersson G, Bertagnoli A, Christensen H, Cuijpers P, Firth J, et al. Towards a consensus around standards for smartphone apps and digital mental health. World Psychiatry 2019 Feb;18(1):97-98 [FREE Full text] [doi: 10.1002/wps.20592] [Medline: 30600619]
39. Smith A. Mashable. 2012. More Smartphone Users Turning to Mobile for Health Info URL: http://mashable.com/2012/11/08/mobile-health-adoption/ [accessed 2019-01-10] [WebCite Cache ID 75QMITzRc]
40. Grynszpan O, Martin JC, Nadel J. Multimedia interfaces for users with high functioning autism: an empirical investigation. Int J Hum-Comput Stud 2008 Aug;66(8):628-639. [doi: 10.1016/j.ijhcs.2008.04.001]
41. Aresti-Bartolome N, Garcia-Zapirain B. Technologies as support tools for persons with autistic spectrum disorder: a systematic review. Int J Environ Res Public Health 2014 Aug 4;11(8):7767-7802 [FREE Full text] [doi: 10.3390/ijerph110807767] [Medline: 25093654]
42. Ploog BO, Scharf A, Nelson D, Brooks PJ. Use of computer-assisted technologies (CAT) to enhance social, communicative, and language development in children with autism spectrum disorders. J Autism Dev Disord 2013 Feb;43(2):301-322. [doi: 10.1007/s10803-012-1571-3] [Medline: 22706582]
43. Bölte S, Golan O, Goodwin MS, Zwaigenbaum L. What can innovative technologies do for autism spectrum disorders? Autism 2010 May;14(3):155-159. [doi: 10.1177/1362361310365028] [Medline: 20603897]
44. Ramdoss S, Lang R, Mulloy A, Franco J, O'Reilly M, Didden R, et al. Use of computer-based interventions to teach communication skills to children with autism spectrum disorders: a systematic review. J Behav Educ 2010 Nov 4;20(1):55-76. [doi: 10.1007/s10864-010-9112-7]
45. Duggan C, Parry G, McMurran M, Davidson K, Dennis J. The recording of adverse events from psychological treatments in clinical trials: evidence from a review of NIHR-funded trials. Trials 2014 Aug 27;15:335 [FREE Full text] [doi: 10.1186/1745-6215-15-335] [Medline: 25158932]
46. Grist R, Cavanagh K. Computerised cognitive behavioural therapy for common mental health disorders, what works, for whom under what circumstances? A systematic review and meta-analysis. J Contemp Psychother 2013 Sep 4;43(4):243-251. [doi: 10.1007/s10879-013-9243-y]
47. Higgins JP, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions. Chichester (UK): John Wiley & Sons; 2011.
48. Baumeister H, Hutter N, Bengel J. Psychological and pharmacological interventions for depression in patients with diabetes mellitus and depression. Cochrane Database Syst Rev 2012 Dec 12;12:CD008381. [doi: 10.1002/14651858.CD008381.pub2] [Medline: 23235661]
49. Gulliver A, Griffiths KM, Christensen H, Brewer JL. A systematic review of help-seeking interventions for depression, anxiety and general psychological distress. BMC Psychiatry 2012 Jul 16;12:81 [FREE Full text] [doi: 10.1186/1471-244X-12-81] [Medline: 22799879]
50. Nugent BM, O'Donnell CM, Epperson CN, Bale TL. Placental H3K27me3 establishes female resilience to prenatal insults. Nat Commun 2018 Jul 2;9(1):2555 [FREE Full text] [doi: 10.1038/s41467-018-04992-1] [Medline: 29967448]
51. NHS England. 2016. Children and Young People's Mental Health Local Transformation Plans - A Summary of Key Themes URL: https://www.england.nhs.uk/mentalhealth/wp-content/uploads/sites/29/2016/08/cyp-mh-ltp.pdf [accessed 2019-01-10] [WebCite Cache ID 75FDHjq8G]
52. Bashiri A, Ghazisaeedi M, Shahmoradi L. The opportunities of virtual reality in the rehabilitation of children with attention deficit hyperactivity disorder: a literature review. Korean J Pediatr 2017 Nov;60(11):337-343 [FREE Full text] [doi: 10.3345/kjp.2017.60.11.337] [Medline: 29234356]
53. Reger GM, Holloway KM, Candy C, Rothbaum BO, Difede J, Rizzo AA, et al. Effectiveness of virtual reality exposure therapy for active duty soldiers in a military mental health clinic. J Trauma Stress 2011 Feb;24(1):93-96. [doi: 10.1002/jts.20574] [Medline: 21294166]
54. Valmaggia LR, Latif L, Kempton MJ, Rus-Calafell M. Virtual reality in the psychological treatment for mental health problems: an systematic review of recent evidence. Psychiatry Res 2016 Feb 28;236:189-195. [doi: 10.1016/j.psychres.2016.01.015] [Medline: 26795129]

55. Torous J, Firth J. The digital placebo effect: mobile mental health meets clinical psychiatry. Lancet Psychiatry 2016 Feb;3(2):100-102. [doi: 10.1016/S2215-0366(15)00565-9] [Medline: 26851322]
56. Carr ME, Anderson A, Moore DW, Evans WH. How should we determine treatment effectiveness with single-case design research for participants with autism spectrum disorder? Rev J Autism Dev Disord 2014 Jul 19;2(1):8-18. [doi: 10.1007/s40489-014-0030-9]
57. Carr M, Moore D, Anderson A. Self-management interventions on students with autism. Except Child 2014 Sep 16;81(1):28-44 [FREE Full text] [doi: 10.1177/0014402914532235]

Abbreviations

ADHD: Attention deficit hyperactivity disorder
ASD: Autism Spectrum Disorder
CBIT: Comprehensive Behavioural Intervention for Tics
CBT: cognitive behavioral therapy
CG: control group
CTD: Chronic Tic Disorder
CYP: children and young people
DSM: Diagnostic and Statistical Manual of Mental Disorder
ER: emotion recognition
F2F: face-to-face
HFASD: high-functioning autism spectrum disorder
ITT: intention-to-treat
JBI: Joanna Briggs Institute
LD: learning disorder
MEDLINE: Medical Literature Analysis and Retrieval System Online
NDD: neurodevelopmental disorder
NHS: National Health Service
NIHR: National Institute for Health Research
OT: occupational therapy
RCT: randomized controlled trial
SMD: standardized mean difference
SUMARI: System for the Unified Management, Assessment and Review of Information
TAU: treatment as usual
TD: tic disorder
TOBY: Therapy Outcomes By You
VMI: visual-motor integration
WLC: waitlist control

Edited by G Eysenbach; submitted 11.03.19; peer-reviewed by A Anderson, C Liang, A Wickersham; comments to author 25.04.19; revised version received 12.06.19; accepted 28.07.19; published 01.11.19.

Please cite as:
Khan K, Hall CL, Davies EB, Hollis C, Glazebrook C
The Effectiveness of Web-Based Interventions Delivered to Children and Young People With Neurodevelopmental Disorders: Systematic Review and Meta-Analysis
J Med Internet Res 2019;21(11):e13478
URL: https://www.jmir.org/2019/11/e13478
doi:10.2196/13478
PMID:31682573

Original Paper

A New Mental Health Mobile App for Well-Being and Stress Reduction in Working Women: Randomized Controlled Trial

Cássia Canha Coelhoso[1], MS; Patricia Renovato Tobo[2*], PhD; Shirley Silva Lacerda[1*], PhD; Alex Heitor Lima[1], BEng; Carla Regina Camara Barrichello[2], BPharm; Edson Amaro Jr[1], PhD; Elisa Harumi Kozasa[1], PhD

[1]Hospital Israelita Albert Einstein, Sao Paulo, Brazil

[2]Natura Cosméticos SA, Cajamar, Brazil

[*]these authors contributed equally

Corresponding Author:
Elisa Harumi Kozasa, PhD
Hospital Israelita Albert Einstein
Av. Albert Einstein, 627/701 bloco A 2ºSS
Sao Paulo, 05601-901
Brazil
Phone: 55 11 99179 9721
Email: ehkozasa@gmail.com

Abstract

Background: Although the availability and use of mobile mental health apps has grown exponentially in recent years, little data are available regarding their efficacy.

Objective: This study aimed to evaluate the effectiveness of an app developed to promote stress management and well-being among working women compared with a control app.

Methods: Female employees at a private hospital were invited to participate in the study via mailing lists and intranet ads. A total of 653 individuals self-enrolled through the website. Eligible participants were randomized between control (n=240) and intervention (n=250) groups. The well-being mobile app provides an 8-week program with 4 classes per week (including a brief theoretical portion and a 15-min guided practice). The active control app also provided 4 assessments per week that encouraged participants to self-observe how they were feeling for 20 min. We also used the app to conduct Web-based questionnaires (10-item Perceived Stress Scale and 5-item World Health Organization Well-Being Index) and ask specific questions to assess subjective levels of stress and well-being at baseline (t_1), midintervention (t_4=4 weeks after t_1) and postintervention (t_8=8 weeks after t_1). Both apps were fully automated without any human involvement. Outcomes from the control and intervention conditions at the 3 time points were analyzed using a repeated measures analysis of variance.

Results: Among the randomized participants (n=490), 185 participants were excluded at the 4-week follow-up and another 79 at the 8-week follow-up because of noncompliance with the experimental protocol. Participants who did not complete t_4 and t_8 assessments were equally distributed between groups (t_4: control group=34.6% [83/240] and intervention group=40.8% [102/250]; P=.16; t_8: control group=29.9% [47/157] and intervention group=21.6% [32/148]; P=.10). Both groups showed a significant increase in general well-being as a function of time ($F_{2,426}$=5.27; P=.006), but only the intervention group presented a significant increase in work-related well-being ($F_{2,426}$=8.92; P<.001), as well as a significant reduction in work-related and overall stress ($F_{2,426}$=5.50; P=.004 and $F_{2,426}$=8.59; P<.001, respectively).

Conclusions: The well-being mobile app was effective in reducing employee stress and improving well-being.

Trial Registration: Clinicaltrials.gov NCT02637414; https://clinicaltrials.gov/ct2/show/NCT02637414.

(J Med Internet Res 2019;21(11):e14269) doi:10.2196/14269

KEYWORDS

stress, psychological; mental health; health promotion; mobile applications; mind-body therapies; meditation; behavioral symptoms; behavioral medicine; psychology; women's health

Introduction

Women's Mental Health

Over the last few decades, studies have been investigating the impact of occupational stress complaints on workers' mental health [1,2]. Currently, there is a pressing need for a greater understanding of gender as a social determinant of health in the work context [3]. According to the American Psychological Association's report *Stress in America: The State of Our Nation* [4], women consistently report higher stress levels than men, are more likely to say they experience symptoms of stress, and have difficulty dealing with it. Several studies indicate that work-related factors may affect women and men differently and suggest that women may be disproportionately affected because of work and family roles [5,6]. Research from the World Economic Forum's Gender Gap study showed that women work nearly an hour longer than men every day, when both paid and unpaid tasks are considered [7]. Therefore, these findings indicate that working women may experience more stress than men and that the sources of stress are related to the expected and actual roles of women in society. Considering the increasing numbers of women in the workforce, it is critical that more resources be directed toward reducing the negative effects of work stress on women's health and well-being.

Meditation and Positive Psychology

Some very successful programs in stress management are based on the principle of Mindfulness—a metacognitive ability, defined by Jon Kabat-Zinn as "the awareness that emerges through paying attention on purpose, in the present moment, and nonjudgmentally to the unfolding experience moment to moment" [8]. The mechanism of action for mindfulness may be explained through a set of 4 interacting components [9]: (1) attention regulation (sustained attention, with returned attention on the main object of focus upon distraction), (2) body awareness (focused attention on subtle bodily sensations to enhance attunement with one's body), (3) emotional regulation (practice of nonjudgmental awareness of one's emotional responses in the moment), and (4) change in self-perspective (detachment from the view of an unchanging self). Although research exploring the effectiveness of online mindfulness-based interventions (MBIs) is still in its infancy, a recent meta-analysis concluded that there is emerging evidence that online MBIs have the potential to improve mental health outcomes, most notably stress [10].

Although the mental health care system has traditionally focused more on treatment of mental disorders than on prevention, it is recognized that mental health is more than just the absence of mental illness. Positive psychology is the study of well-being, engagement, and optimal functioning, fitting well with the World Health Organization's (WHO) definition of mental health: "a state of well-being in which the individual realizes his or her own abilities, can cope with the normal stresses of life, can work productively and fruitfully, and is able to make a contribution to his or her community" [11]. The PERMA model proposed by Seligman [12] stipulates that happiness and psychological well-being is made up of 5 core components: positive emotions, engagement (similar to the concept of mindfulness), positive relationships, meaning, and accomplishments. Well-being has been shown to reduce the risk of developing mental health alterations or disorders [13,14]. Results from a meta-analysis of 51 positive psychology interventions with more than 4000 individuals revealed that such interventions do indeed significantly enhance well-being [15]. A number of studies conducted with adult [16] and younger [17] populations have shown improvements in well-being through the use of online positive psychology interventions. Technology-based self-care programs allow individuals to circumvent some of the limitations associated with traditional support methods, such as time, resources, flexibility, accessibility, and availability [18]. Indeed, this is in line with WHO directives recommending "the promotion of self-care, for instance, through the use of electronic and mobile health technologies" [19].

Mobile Health

Mobile health (mHealth), which promotes the use of wireless technologies in health care, is one of the fastest growing fields within electronic health (eHealth) [20]. According to the *Digital 2019 Q2 Global Digital Statshot* report, 5.11 billion people own a mobile phone, which represents 66% of the world's population [21]. About internet and mobile use in Brazil, the recently launched *Digital in 2019* report [22] showed that more than 149 million out of the country's nearly 212 million inhabitants are active internet users and that there are 215.2 million mobile connections in Brazil, which represents a penetration of 102%. Moreover, the report stated that 66% of all Brazilians are mobile internet users and that there is an average use of 34 mobile apps per month per smartphone in the country. According to the IQVIA Institute, there are currently more than 318,000 health apps worldwide, nearly double the number of apps available in 2015, with more than 200 new health apps being added to app stores every day [23]. Even with this substantial increase over the previous report [24], consumer Digital Health apps targeting wellness management still account for the majority of health apps (60%), and mental health remains the largest focus for disease-specific mobile apps (28%) [23].

The use of a mobile device makes it possible to intervene and interact with the participants within the context (ie, their work environment) and during moments of their daily life, which is a form of intervention called ecological momentary intervention (EMI). The most reliable method for investigating real-world emotion is experience sampling, which involves contacting people as they engage in their everyday activities and asking them to report their thoughts, feelings, and actions at that moment [25]. Therefore, app-based EMIs offer a versatile, multifaceted, and interactive way of promoting training, mindfulness, self-awareness, motivation, and environmental awareness within the context of everyday life. The near ubiquitous use of smartphone apps provides a vehicle for making EMI a widespread and effective way of promoting positive change in a large nonclinical, nontherapeutic population.

Although smartphones and mobile apps seem to be ideal tools in many aspects for providing instantly accessible interventions to promote health, the content of these apps must be based on systematic research, as without such content, any observed effects could be placebo or even harmful. Even though the

availability and use of mental health mobile apps have grown exponentially in recent years, few have been thoroughly tested to provide robust scientific evidence regarding their efficacy in modulating behavior or promoting health [26-28].

Objectives

From these considerations, we developed 2 apps and hypothesized that a well-being mobile app, composed of an 8-week program with 4 lessons per week—based on relaxation training, breathing techniques, guided meditation, and positive psychology principles—could be effective in reducing stress and increasing well-being among working women when compared with an active control app designed to encourage self-observation and evaluation of subjective levels of stress and well-being, for the same period and weekly frequency. Here, we report the results of a randomized controlled trial in which we compared a well-being mobile app and an active control app on improving well-being and reducing stress in a group of women working in a private hospital in São Paulo, Brazil. In short, the aim of this study was to evaluate whether it is possible to improve psychological health by promoting stress management and well-being using a mobile app.

Methods

Trial Design

The experimental design of this study and the format of the manuscript followed the Consolidated Standards of Reporting Trials (CONSORT) statement for reporting randomized controlled trials [29], its extension for nonpharmacological trials [30], and the CONSORT-EHEALTH checklist [31]. See Multimedia Appendix 1 for CONSORT-EHEALTH checklist (V 1.6.1). The trial was registered at ClinicalTrials.gov, number NCT02637414 (Flourishing App: Evaluation of the Effectiveness of a Well-being App for Mobile Devices), on December 11, 2015. This study was a 2-arm randomized controlled trial conducted in Brazil. Participants were randomized using a 1:1 allocation ratio to 1 of the 2 parallel groups: (1) well-being mobile app based on relaxation, breathing, meditation, and positive psychology principles and (2) control app, containing only instructions to self-observation for 20 min and recording of subjective levels of stress and well-being. Both apps were developed in Brazilian Portuguese language and were fully automated without any human involvement. Longitudinal assessments were conducted at baseline, midintervention (4 weeks after baseline), and postintervention (8 weeks after baseline).

Participants

We conducted this study at a large private tertiary care hospital in São Paulo, Brazil, from June 2016 to May 2017. We used mailing lists and intranet ads to invite hospital staff to participate in the study. Individuals were included if they met the following inclusion criteria: women aged between 20 and 60 years who had completed high school, owned a mobile device with either an iOS or Android operating system, and were available to participate in the 8-week training program. Potential participants first provided demographic data through an initial registration form, and those who were selected were then randomly assigned to 1 of the 2 groups (control or intervention). After that, they were given access to 1 of the 2 apps and a tutorial with instructions for its use. Excluded participants were informed via phone or email.

Interventions

Intervention Condition

The well-being mobile app consists of an 8-week program divided into 2 4-week modules. Participants had to attend 4 classes per week; if they did not meet this deadline, they were given 4 more days to complete the classes, after which the app was blocked. Each class contained a brief theoretical portion and a 15-min guided practice. Twice a week, participants were asked to write down 1 good thing they had experienced and 1 good deed they had performed in a gratitude journal. They were told to do the activities whenever and wherever it felt most appropriate. Some screenshots of this app are provided in Multimedia Appendix 2.

The well-being mobile app was designed to handle psychological stress based on relaxation training, breathing techniques, meditation (such as mindfulness, loving, kindness, and empathetic joy), and positive psychology principles. The major aim of relaxation training and the use of breathing techniques is to reduce chronic stress and enhance well-being by eliciting a relaxation response, defined as a physical state characterized by decreased arousal of the sympathetic nervous system and the opposite of the body's stress response to perceived threats [32]. An emerging approach to increase awareness and respond skillfully to mental processes that contribute to emotional distress and maladaptive behavior is based on mindfulness meditation, a process of maintaining a moment-by-moment awareness of our thoughts, feelings, bodily sensations, and surrounding environment, without judging them [8]. Finally, positive psychology activities that aim to cultivate positive feelings, behaviors, or cognitions such as practicing optimistic thinking, expressing gratitude, practicing kindness, and replaying positive experiences are another promising approach to increase psychological well-being [33,34].

During the first 2 weeks, the theoretical content dealt with physical and psychological consequences of chronic stress and the use of breathing techniques as a way to reduce it [32,35]. Guided practice was a technique known as body scan, which involves a gradual sweeping of attention through the entire body, focusing noncritically on any sensation or feeling in body regions and using periodic suggestions of breath awareness and relaxation [36]. The third and fourth weeks focused on mindful breathing to cultivate present moment awareness and meditation, respectively [8]. Guided practice during these 2 weeks involved meditation based on breath counting, using the breath as an anchor when the mind starts to wander [37]. The fifth week presented information on cultivating positive emotions and its impact on one's health, life, and interpersonal relationships [12]. It included a guided loving kindness meditation, which is a method for developing the heartfelt yearning that all may find happiness [38]. During the sixth week, the theoretical content focused on the influence of positive and negative thoughts on different emotional states [12], and the guided exercise trained focusing attention and increasing awareness of one's own

repertoire of emotions, thoughts, and sensations [37]. During the seventh (second to last) week, the concept of empathy was explored [12], and the practice was a guided empathetic joy meditation aimed to train the manifestation of contentment through one's own and others' happiness [38]. The focus of the final week was gratitude and the different ways to cultivate it in daily routine [12], and the guided practice was mindfulness of breathing to focus on what participants were grateful for in the present moment [36]. Finally, all the practices taught during the stress management and wellness promotion program were reviewed. Toward the end of the intervention, participants were encouraged to keep practicing and to apply what they learned to everyday activities.

Control Condition

The control app had the same features as the well-being mobile app, including a menu, a tutorial, a profile page, evaluations, pop-up messages, and push notifications. Some screenshots of this app are provided in Multimedia Appendix 3. This active control condition followed the same period as the intervention. As in the intervention condition, participants had to answer 4 assessments per week. Each assessment was composed of a pre- and postevaluation separated by a 20-min interval. During this period, the control group received the exactly following instruction "Within 20 minutes, you will be invited to respond to these questions again. During this period, try to observe yourself and see how you are feeling," whereas the intervention group carried out the proposed activities in the well-being mobile app. Due to this, we decided to call the active control condition *Monitoring of Perceptions*.

This approach was chosen as an active control for involving a mindfulness training (being aware of their perceptions) and is significantly better than the waiting list condition, which does not consider nonspecific effects of training [39], such as received attention and demand characteristics, or the possibility of a *digital placebo effect* (ie, placebo-like effects seen from mobile health interventions, such as smartphone apps) [40]. Recent technological advances have led to the design of well-being interventions that include adequate experimental controls [41].

Adherence Rates

The adherence rates were collected for all participants by accessing their accounts. The steps of users' navigation flow in both apps were controlled by the following 3 steps: (1) completion of the preevaluation, (2) class (intervention condition) or monitoring of perceptions (control condition), and (3) completion of the postevaluation. We had a new record in the database every time the user completed any evaluation, and we considered an activity as concluded if the user filled the respective postevaluation form. Adherence was based on the number of classes or assessments that the participant concluded during the first 4 weeks and at the end of the 8-week period of the study. Moreover, 1 push notification per day and 2 emails per week were sent to participants to engage them.

Outcomes: Data Collection

Primary outcome measures were taken at baseline (t_1), midintervention (t_4=4 weeks after t_1), and postintervention (t_8=8 weeks after t_1). We assessed stress perception using the Brazilian Portuguese version of the 10-item perceived stress scale (PSS-10) [42,43] and subjective well-being with the Brazilian Portuguese version of the 5-item World Health Organization Well-Being Index (WHO-5) [44,45]. These Web-based questionnaires were applied using Google Forms through a link sent to participants via email. We also used an in-app questionnaire to assess subjective symptoms of stress and well-being at work and overall during the previous 30-day period. Moreover, within the app, secondary measures were gathered 4 times a week during 8 weeks, before and after the period of each class (intervention condition) or monitoring of perceptions (control condition) by assessing current subjective symptoms of stress and well-being.

Primary Outcome Measures

10-Item Perceived Stress Scale

This scale was developed by Cohen et al [46] to assess the degree to which individuals perceive their life situations as stressful. The perceived stress assessed by PSS corresponds to the issue of the cognitive appraisal process, when a "situation is appraised both as threatening or otherwise demanding and as taxing or exceeding the coping resources of the person." In summary, the PSS items check how much respondents consider their lives unpredictable, uncontrollable, and overloaded; it also includes the number of items inquiring about current levels of experienced stress. PSS is a general scale as the content of its questions is not specific to any subpopulation or age group; however, it targets participants who have completed junior high school. The 10 questions address the frequency of participants' feelings and thoughts about events and situations that occurred during the previous 30 days. A total of 6 questions are negative (1, 2, 3, 6, 9, and 10) and 4 are positive (4, 5, 7, and 8). Each question is rated on a 5-point Likert scale from 0 (never) to 4 (very often). To calculate the total score, the 4 positive items are reverse scored and summed across all scale items. Total scores may range from 0 to 40, and higher scores indicate higher levels of perceived stress. As PSS is not a diagnostic measure and there is no official cut-off available, recent studies [47,48] used 1 standard deviation (SD 6.2) above the mean PSS-10 of 15.3 in a large working population [49] as a cut-off value to choose participants with elevated stress levels (PSS-10≥22). Heber et al [47] also calculated the clinically significant change, according to the method developed by Jacobson and Truax [50], and determined that participants changed reliably if their PSS-10 score differed by more than ±5.16 points between assessments. Finally, Ebert et al [48] considered the symptom-free status, defined at PSS-10 <17.70, as the effect outcome for the cost-effectiveness analysis of their intervention. In our sample, the Cronbach alpha for the internal reliability of the PSS-10 was .82, which is similar to that observed in the original study (alpha=.78) [42] and in the Brazilian Portuguese version of the PSS-10 (alpha=.87) [43].

5-Item World Health Organization Well-Being Index

WHO-5 is a short questionnaire consisting of 5 simple, noninvasive, and positively phrased items for measuring subjective well-being. Participants rate each of the 5 statements on a 6-point Likert scale, from 5 (all of the time) to 0 (at no time) to indicate how they felt during the last 2 weeks. The raw

score theoretically ranges from 0 (absence of well-being) to 25 (maximal well-being), but it is recommended that one multiply the result by 4 to translate it to a percentage scale from 0 to 100, with higher scores meaning better subjective well-being. A score below 50% suggests poor emotional well-being [51], and the threshold for a clinically relevant change is considered to be 10% on standardized percentage scores [52]. A recent systematic review [52] showed that the WHO-5 is a helpful tool for clinical practice and research to evaluate well-being over time or to compare well-being between groups. The authors recommend using the general population's mean score as a reference when using the WHO-5 as an outcome measure in clinical trials. In our sample, the Cronbach alpha was .88, which indicates a good internal validity. Similar internal consistency was found in the Brazilian Portuguese version of the WHO-5 (alpha=.83) [45].

Subjective Symptoms of Stress and Well-Being During the Last Month

We assessed subjective symptoms of stress and well-being at work and in general during the previous 30 days using a sliding percentage scale from 0 (absent) to 100 (maximal). Participants answered the following 4 simple questions: (1) What was your level of stress at work during the last month? (2) In general, what was your level of stress during the last month? (3) What was your level of well-being at work during the last month? and (4) In general, what was your level of well-being during the last month?

Secondary Outcome Measures

Subjective Symptoms of Stress and Well-Being at the Moment

We evaluated subjective symptoms of stress and well-being before and after the period of each class (intervention condition) or monitoring of perceptions (control condition) using a sliding percentage scale from 0 (absent) to 100 (maximal). At this time, 2 simple questions were presented within the app: (1) What is your level of stress at this moment? and (2) What is your level of well-being at this moment?

Sample Size

The sample size was calculated according to a confidence interval of 0.95, a sampling error of 0.05, and a power effect of 0.8. From these data, a sample size calculation was conducted, and a minimum requirement of 100 participants in each group was determined.

Recruitment, Randomization, and Blinding

Potential participants were invited to participate by email and hospital intranet ads. Candidates were asked to access the study's website [53], read the information package, and fill out an online registration form. Participants who met the inclusion criteria were randomly assigned to 1 of the 2 trial arms. A research assistant with no clinical involvement in the trial used the Microsoft Excel RANDBETWEEN function to randomly assign either 0 or 1 to each participant, corresponding to each experimental arm. The same research assistant generated and distributed unique access codes. The participants were aware of their allocated study arm, whereas outcome assessors and data analysts remained blind to group allocations until the completion of the study.

Statistical Analysis

Postintervention analyses involved a modified intention-to-treat design excluding participants who did not adequately adhere to the protocol. We developed the app to be a stress management and wellness promotion training program. As courses or trainings usually require a minimum frequency of 75% so that the participant does not fail because of absences, we decided to use this cut-off, even if it was somewhat stringent for mobile apps. Thus, only volunteers who completed 12 of the 16 activities during the first 4 weeks, completing 75% of module I, were included in the sociodemographic analysis, as well as the analysis comparing baseline with mid-intervention scores. Likewise, only volunteers who completed 23 of the 31 activities of the following 4 weeks, completing 75% of module II, were included in the analyses.

Statistical analyses were conducted using IBM SPSS Statistics for Windows, version 24.0 (IBM Corp). Descriptive analyses were used for sociodemographic data to report absolute and relative frequencies as well as means, standard deviations, and medians. The Kolmogorov-Smirnov test was used to evaluate data normality. The Pearson chi-square test was used to compare categorical variables between groups at baseline, whereas Student *t* tests (or the nonparametric Wilcoxon-Mann-Whitney test, WMW) were used to compare the continuous variables. A repeated measures analysis of variance (ANOVA) was used to evaluate primary outcomes, considering 2 conditions (control and intervention) and 3 distinct periods (baseline, midintervention, and postintervention). The time × group interaction effects were assessed to investigate differences regarding the magnitude of change on dependent variables between the experimental and control groups. For the secondary outcome measures (ie, assessments made before and after each activity), we used the WMW test to compare stress and well-being level variations (ΔSL and ΔWBL, respectively) between groups. The significance threshold was set at .05.

Ethics and Consent

Concerning data confidentiality, special efforts were made to secure online data collection and storage. Data collected using Google Forms were only accessed by the study's first author. When these data were downloaded, other authors had access to the relevant data files. Data collected via the project website and within the app were stored on Amazon Relational Database Service. The platform used was an AWS Elastic Beanstalk, and its security criteria are available on the Web page [54]. The server stores data in an unencrypted database, but access is granted through authentication by inserting a log-in, password, and security token. All data were saved in both .xlsx and .sav formats. These files were stored in a password-protected folder on a secure server and were only available to the research team. In accordance with the Hospital Israelita Albert Einstein data retention policy, these data will be retained for 5 years after the end of the study and anonymized by replacing participant names with subject ID numbers.

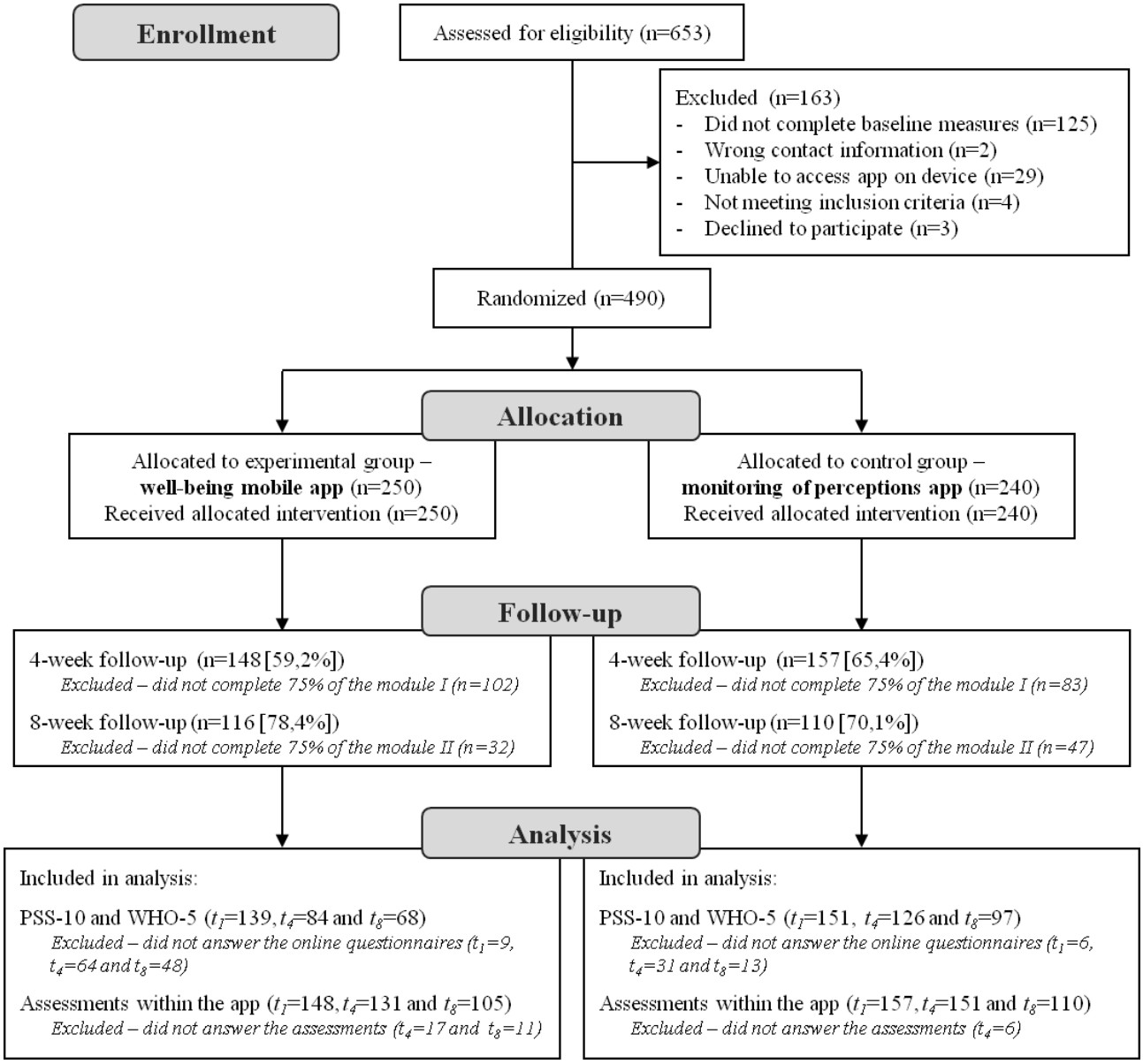
The study protocol and consent procedure were approved by the Ethics Committee of the Instituto Israelita de Ensino e Pesquisa Albert Einstein—Brazil (nº 1.469.020). Study participants were given adequate information before agreeing to participate and freely providing informed consent. Documentation of written consent was not mandatory for this study because an online consent form was used (available on the study's website) and participants could only access the registration form if they accepted and signed the consent form [53]. It was clearly communicated that their participation was voluntary and without monetary compensation. Participants were reminded that they were free to withdraw at any time, without penalty or need for explanation, and that their data would be stored securely and anonymously. Control participants were offered access to the well-being mobile app upon study completion. Although there were no reported risks associated with the stress management and wellness promotion program or similar online interventions, the questionnaires and activities could cause discomfort for some people. Given that there were no adverse effects associated with using the app, we had no reason to discontinue the intervention for individuals who chose to continue.

Results

Participant Flow

In September 2016, 653 individuals were enrolled in the study through the website and were then screened for eligibility. Among these potential participants, 163 were excluded (because they did not complete baseline measures or were unable to access the app) and the remaining 490 were randomized (see Figure 1 for a flow diagram). We excluded an additional 185 participants at the 4-week follow-up and another 79 at the 8-week follow-up because of noncompliance with the experimental protocol (ie, completed less than 75% of each module). Participants who did not complete t_4 and t_8 assessments were equally distributed between groups (t_4: control group=34.6% [83/240] and intervention group=40.8% [102/250]; X^2_1=2.0; P=.16; effect size for chi-square test φ=0.064; t_8: control group=29.9% [47/157] and intervention group=21.6% [32/148]; X^2_1=2.7; P=.10; φ=0.095).

Figure 1. Consolidated Standards of Reporting Trials flow diagram. PSS-10: 10-item Perceived Stress Scale; WHO-5: 5-item World Health Organization Well-Being Index.



Baseline Data

Table 1 and Figure 2 show the baseline sociodemographic and outcome score data for both groups, respectively. Overall, the mean age of the participants was 34.6 (SD 7.62) years; 53% (n=161) were married or lived with a domestic partner; 45% (n=137) had a graduate degree; and 65% (n=197) were professionals other than nurses, physicians, or professionals in leadership positions. On average, PSS-10 scores were 22.73 (SD 6.69), and WHO-5 scores were 10.75 (SD 5.14). As seen in Table 1 and Figure 2, baseline characteristics were equivalent. See Multimedia Appendix 4 for mean (SD) and median (p25-p75) values.

Table 1. Baseline sociodemographic data by experimental group.

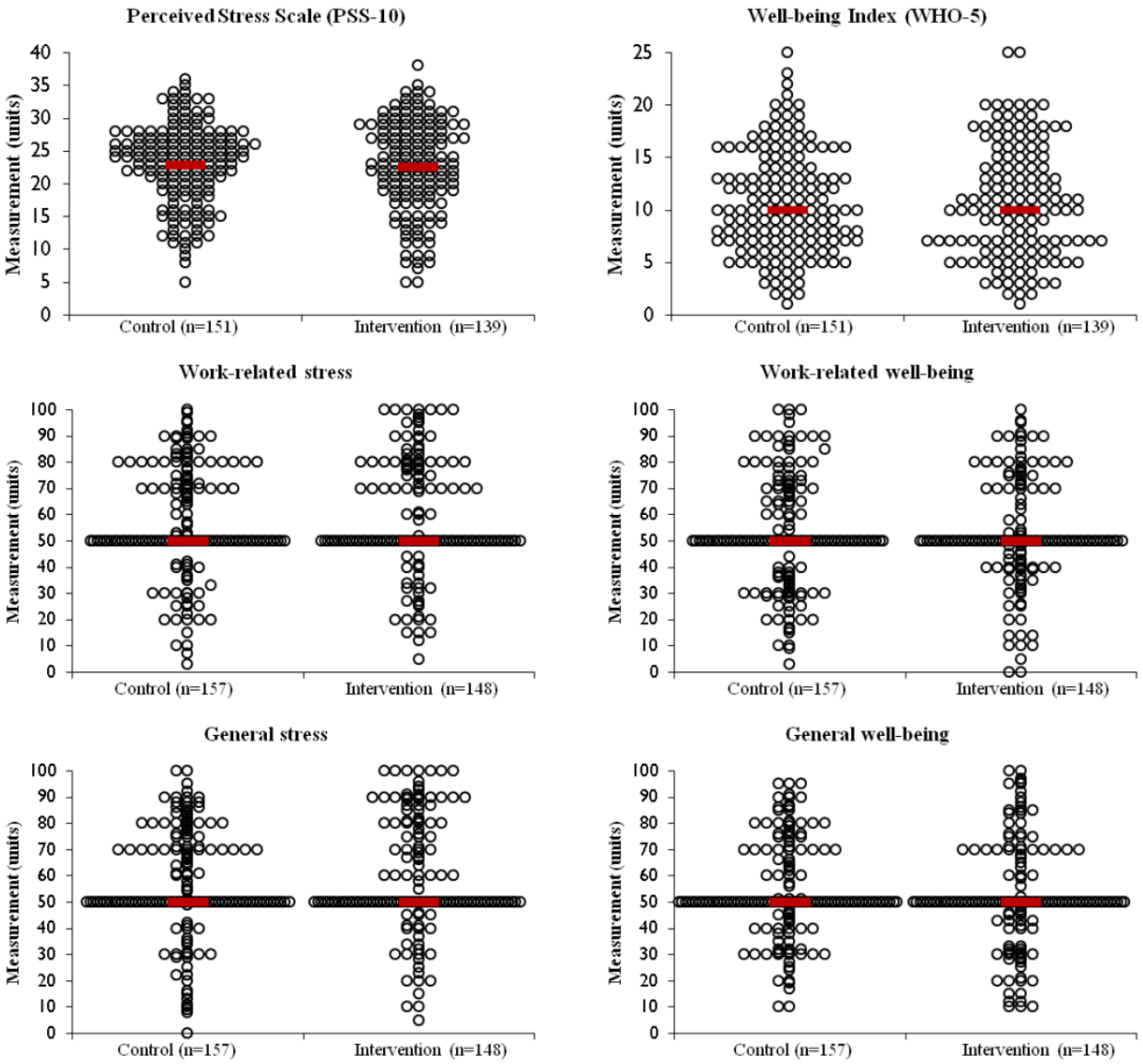
Characteristics[a]	Control (n=157)	Intervention (n=148)	*P* value
Age (years), mean (SD)	33.8 (7.47)	35.4 (7.73)	.07
Children, median (p25-p75)[b]	0 (0-1)	1 (0-1)	.75
Profession, n (%)			.77
Assistant	0 (0.0)	1 (0.7)	—[c]
Analyst	1 (0.6)	1 (0.7)	—
Leadership coordinator	10 (6.4)	8 (5.4)	—
Leadership manager	1 (0.6)	0 (0.0)	—
Professional	105 (66.9)	92 (62.2)	—
Nursing	33 (21.0)	39 (26.4)	—
Physician	7 (4.5)	7 (4.7)	—
Marital status, n (%)			.68
Single	61 (38.9)	52 (35.1)	—
Married/cohabiting	78 (49.7)	83 (56.1)	—
Separated/divorced	16 (10.2)	12 (8.1)	—
Others	2 (1.3)	1 (0.7)	—
Educational level, n (%)			.40
Complete high school	30 (19.1)	28 (18.9)	—
Incomplete higher education	20 (12.7)	25 (16.9)	—
Complete higher education	39 (24.8)	26 (17.6)	—
Postgraduate degree	68 (43.3)	69 (46.6)	—

[a]Statistical test results comparing age (t_{303}=−1.803; Cohen d=0.21), number of children (Mann-Whitney U test, U=11388; effect size correlation for Mann-Whitney U test, r=−0.018), profession (X^{-2}_{6}=3.3; effect size Cramer V=0.104), marital status (X^{-2}_{3}=1.5; V=0.070) and educational level (X^{-2}_{3}=2.9; V=0.099) for the control and intervention groups.

[b]Data are presented as medians (25th-75th percentile).

[c]Not applicable.

Figure 2. Baseline outcome score data by experimental group. Scatter plots showing bivariate analyses of 10-item Perceived Stress Scale (t_{288}=0.414; *P*=.68; *d*=0.045), 5-item World Health Organization Well-Being Index (*U*=10461; *P*=.96; *r*=−0.002), work-related stress (*U*=11078; *P*=.47; *r*=−0.041), general stress (*U*=11274; *P*=.64; *r*=−0.026), work-related well-being (*U*=11487; *P*=.86; *r*=−0.010), and general well-being (*U*=11060; *P*=.45; *r*=−0.043) for the control and intervention groups. Open circles show data for each participant. Red lines show group medians, except for PSS-10, which shows group means. PSS-10: 10-item Perceived Stress Scale; WHO-5: 5-item World Health Organization Well-Being Index.



Outcomes

Preintervention × Midintervention Analysis

Efficacy results from the first 4 weeks of intervention, using a per-protocol approach, are presented in Figure 3. See Multimedia Appendix 5 for descriptive data related to repeated measures analysis of variance. Across groups, PSS-10 scores significantly decreased from preintervention (sample mean, $\bar{x}$=22.35; SE=0.48) to midintervention ($\bar{x}$=19.65; SE=0.47; $F_{1,208}$=41.9; *P*<.001; η_p^2=0.168). A time × group interaction ($F_{1,208}$=9.48; *P*=.002; η_p^2=0.044) indicated that the intervention group experienced significant decreases in perceived stress compared with the control condition.

WHO-5 scores increased for all participants as a function of time ($F_{1,208}$=35.4; *P*<.001; η_p^2=0.146). A time × group interaction was also observed ($F_{1,208}$=8.54; *P*=.004; η_p^2=0.168) with participants in the intervention condition showing a greater increase on the well-being index from preintervention to midintervention than participants in the control group.

Work-related stress did not change as a function of time alone ($F_{1, 280}$=2.54; *P*=.11; η_p^2=0.009) but changed as a function of time × group ($F_{1,280}$=6.34; *P*=.01; η_p^2=0.022), such that people in the intervention condition reported greater improvements in work-related stress from preintervention to midintervention compared with participants in the control group.

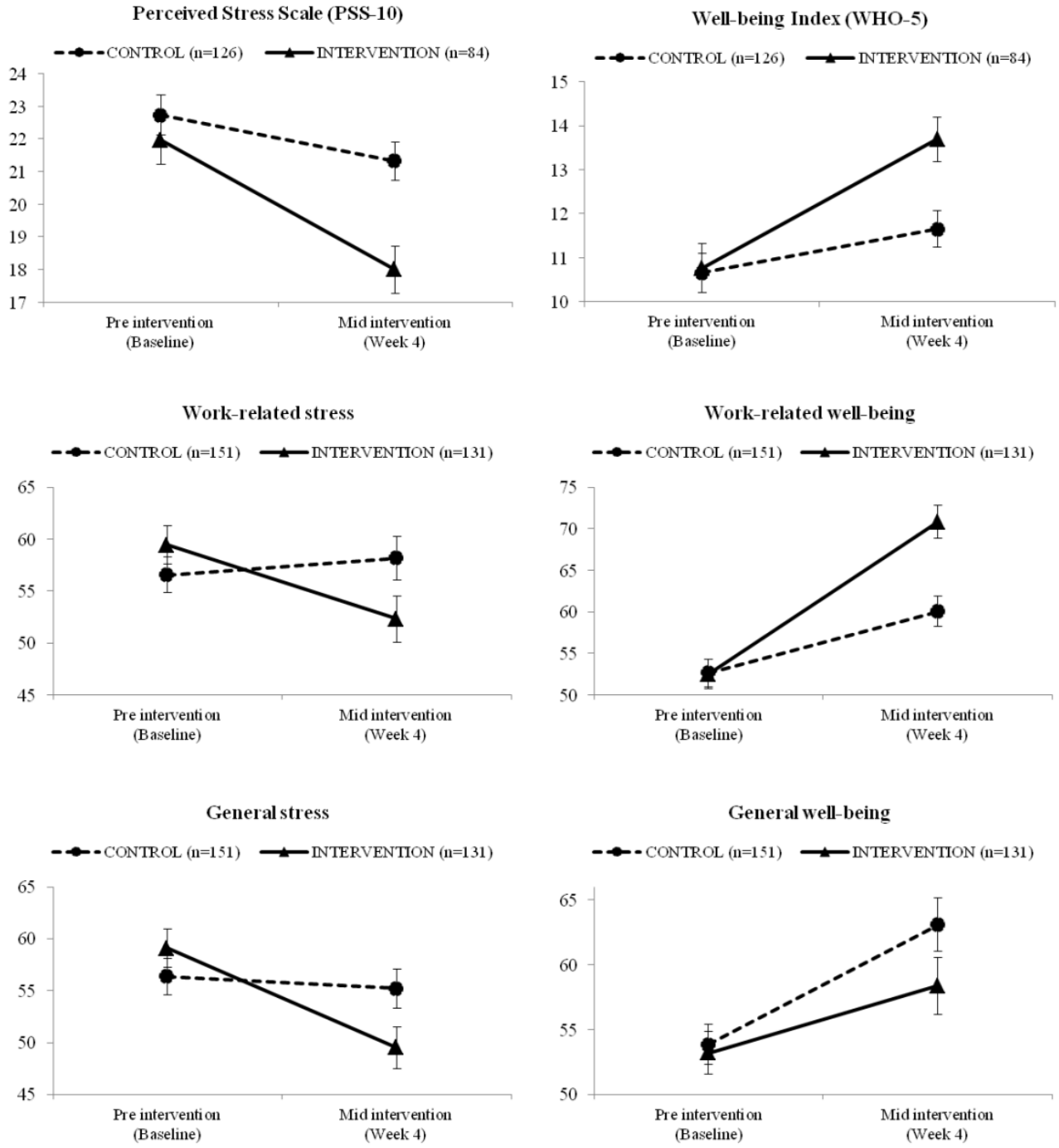
General stress showed a significant time effect ($F_{1,280}$=10.4; *P*=.001; η_p^2=0.036), with lower general stress ratings from preintervention ($\bar{x}$=57.71; SE=1.26) to midintervention ($\bar{x}$=52.37; SE=1.38). A significant time × group interaction

emerged ($F_{1,280}$=6.46; P=.01; η_p^2=0.023), such that participants in the intervention condition experienced greater reductions in general stress than those in the control group.

Work-related well-being significantly increased across participants from preintervention ($\bar{x}$=52.57; SE=1.22) to midintervention ($\bar{x}$=65.43; SE=1.34; $F_{1,280}$=67.2; P<.001; η_p^2=0.194), and there was a significant time × group interaction ($F_{1,280}$=12.1; P=.001; η_p^2=0.041), showing a greater benefit for individuals in the intervention group.

General well-being increased as a function of time from preintervention ($\bar{x}$=53.53; SE=1.12) to midintervention ($\bar{x}$=60.72; SE=1.52; $F_{1,280}$=18.5; P<.001; η_p^2=0.062), but no time × group interaction was observed.

Figure 3. Plot means with standard error bars showing all outcome data for the control (dashed black line) and intervention (solid black line) groups at pre intervention and mid intervention. Time × group interaction at 10-item Perceived Stress Scale ($F_{1,208}$=9.48; P=.002; η_p^2=.044), 5-item World Health Organization Well-Being Index ($F_{1,208}$=8.54; P=.004; η_p^2=.168), work-related stress ($F_{1,208}$=6.34; P=.01; η_p^2=.022), work-related well-being ($F_{1,208}$=12.1; P=.001; η_p^2=.041), general stress ($F_{1,208}$=6.46; P=.01; η_p^2=.023), and general well-being ($F_{1,208}$=1.48; P=.22; η_p^2=.005). PSS-10: 10-item Perceived Stress Scale; WHO-5: 5-item World Health Organization Well-Being Index.



Preintervention × Midintervention × Postintervention Analysis

Intervention efficacy results using a per-protocol approach are shown in Figure 4. See Multimedia Appendix 6 for descriptive data related to repeated measures analysis of variance. Analysis of PSS-10 scores revealed a time effect ($F_{2,326}$=45.0; P<.001; η_p^2=0.216), and pairwise comparisons indicated significant declines in PSS-10 scores from preintervention ($\bar{x}$=22.27; SE=0.56) to midintervention ($\bar{x}$=19.60; SE=0.53), and again from midintervention to postintervention ($\bar{x}$=17.88; SE=0.54). A time × group interaction ($F_{2,326}$=7.19; P=.001; η_p^2=0.042) indicated that the intervention group produced significant decreases in perceived stress when compared with the control group.

WHO-5 scores increased across all participants as a function of time ($F_{2,326}$=33.4; P<.001; η_p^2=0.170), and pairwise comparisons indicated significant increases in WHO-5 scores from preintervention ($\bar{x}$=10.69; SE=0.41) to midintervention ($\bar{x}$=12.72; SE=0.37), and again from midintervention to postintervention ($\bar{x}$=13.53; SE=0.37). A time × group interaction was also observed ($F_{2,326}$=7.97; P<.001; η_p^2=0.047), with participants in the intervention condition showing a greater increase in well-being than participants in the control group.

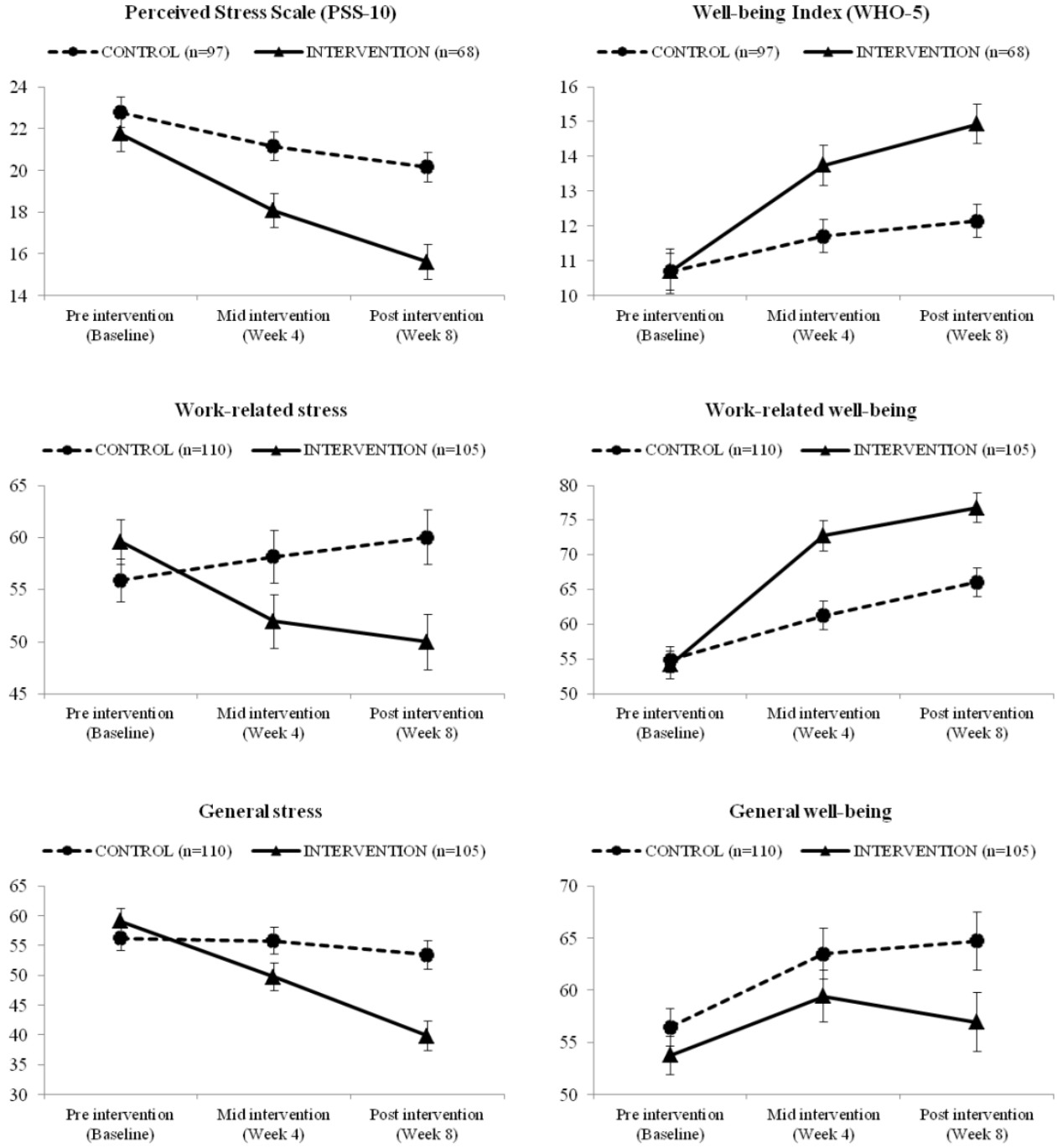
Work-related stress did not change as a function of time alone ($F_{2,426}$=1.09; P=.34; η_p^2=0.005) but changed as a function of time × group ($F_{2,426}$=5.50; P=.004; η_p^2=0.025), such that people in the intervention condition reported greater improvements in work-related stress compared with those in the control group.

General stress showed a significant time effect ($F_{2,426}$=15.3; P<.001; η_p^2=0.067), and pairwise comparisons indicated significant decreases in general stress ratings from preintervention ($\bar{x}$=57.63; SE=1.49) to midintervention ($\bar{x}$=52.72; SE=1.61), and again from midintervention to postintervention ($\bar{x}$=46.60; SE=1.73). A significant time × group interaction was observed ($F_{2,426}$=8.59; P<.001; η_p^2=0.039), such that participants in the intervention condition experienced greater reductions in general stress than those in the control condition.

Work-related well-being significantly increased across participants from preintervention ($\bar{x}$=54.55; SE=1.36) to midintervention ($\bar{x}$=67.03; SE=1.51) and again from midintervention to postintervention ($\bar{x}$=71.43; SE=1.47; $F_{2,426}$=58.5; P<.001; η_p^2=0.215), and there was a significant time × group interaction ($F_{2,426}$=8.92; P<.001; η_p^2=0.040), showing a greater benefit for individuals in the intervention group.

General well-being increased as a function of time ($F_{2,426}$=5.27; P=.006; η_p^2=0.024), and pairwise comparisons indicated significant increases in general well-being ratings from preintervention ($\bar{x}$=55.08; SE=1.27) to midintervention ($\bar{x}$=61.45; SE=1.75) and to postintervention ($\bar{x}$=60.82; SE=1.98). Unlike the other variables analyzed thus far, no differences were observed between mid- and postintervention measurements for general well-being. There was also no time × group interaction.

Figure 4. Plot means with standard error bars showing all outcome data for the control (dashed black line) and intervention (solid black line) groups at preintervention, midintervention, and postintervention. A time × group interaction was observed at 10-item Perceived Stress Scale ($F_{2,326}$=7.19; P=.001; η_p^2=.042), 5-item World Health Organization Well-Being Index ($F_{2,326}$=7.97; P<.001; η_p^2=.047), work-related stress ($F_{2,426}$=5.50; P=.004; η_p^2=.025), work-related well-being ($F_{2,426}$=8.92; P<.001; η_p^2=.040), general stress ($F_{2,426}$=8.59; P<.001; η_p^2=.039), and general well-being ($F_{2,426}$=0.74; P=.47; η_p^2=.003). PSS-10: 10-item Perceived Stress Scale; WHO-5: 5-item World Health Organization Well-Being Index.



Changes in Subjective Symptoms of Stress and Well-Being After the Proposed Daily Activities

For secondary outcome evaluations, we compared stress and well-being level variations (ΔSL and ΔWBL, respectively) between groups after completion of 75% of the classes in module I. As expected, participants using the well-being mobile app reported a significantly greater reduction in stress levels (Table 2) and significantly greater increases in well-being levels (Table 3) than those using the active control app.

Table 2. Between-group comparisons of stress level variations after each class.

Class number	Control		Intervention		Significance		
	n	Median (p25-p75)[a]	n	Median (p25-p75)	*U* [b]	*P* value	*r* [c]
Class 1	126	0.00 (−15.0-4.25)	148	−15.0 (−30.0-0.00)	5897	<.001	−0.309
Class 2	90	0.00 (−9.00-9.25)	148	−12.5 (−27.7-0.00)	3675	<.001	−0.377
Class 3	81	0.00 (−20.0-4.00)	148	−11.5 (−26.7-0.00)	4671	.006	−0.183
Class 4	87	0.00 (−13.0-3.00)	148	−9.50 (−21.0-0.00)	4667	<.001	−0.231
Class 5	83	0.00 (−9.00-5.00)	148	−9.00 (−26.0-0.00)	3352	<.001	−0.378
Class 6	87	0.00 (−13.0-0.00)	148	−10.0 (−30.7-0.00)	4957	.003	−0.193
Class 7	88	0.00 (−6.00-0.75)	148	−8.00 (−21.0-0.00)	3883	<.001	−0.340
Class 8	88	0.00 (−7.75-2.50)	148	−6.00 (−29.0-0.00)	4464	<.001	−0.264
Class 9	98	0.00 (−4.00-3.00)	148	−10.0 (−26.7-0.00)	3945	<.001	−0.389
Class 10	103	0.00 (−9.00-4.00)	148	−8.00 (−20.7-0.00)	5153	<.001	−0.278
Class 11	95	0.00 (−7.00-4.00)	148	−4.50 (−13.7-0.00)	4961	<.001	−0.251
Class 12	97	0.00 (−10.0-1.00)	148	−7.00 (−17.7-0.00)	5170	<.001	−0.238

[a]Data are presented as medians (25th-75th percentile).

[b]Mann-Whitney *U* test.

[c]Effect size correlation for Mann-Whitney *U* test.

Table 3. Between-group comparisons of well-being level variations after each class.

Class number	Control		Intervention		Significance		
	n	Median (p25-p75)[a]	n	Median (p25-p75)	*U* [b]	*P* value	*r* [c]
Class 1	126	1.50 (−5.25-20.0)	148	20.0 (5.25-40.0)	5108	<.001	−0.390
Class 2	90	0.00 (−2.00-10.0)	148	18.0 (4.00-30.7)	3165	<.001	−0.441
Class 3	81	3.00 (0.00-14.0)	148	13.5 (2.00-30.0)	3821	<.001	−0.300
Class 4	87	0.00 (−2.00-13.0)	148	13.0 (0.00-29.2)	4105	<.001	−0.303
Class 5	83	0.00 (−6.00-10.0)	148	11.0 (1.25-25.7)	3163	<.001	−0.404
Class 6	87	1.00 (0.00-10.0)	148	10.0 (0.00-29.7)	4335	<.001	−0.274
Class 7	88	0.00 (−1.00-5.00)	148	13.0 (1.00-29.0)	3057	<.001	−0.447
Class 8	88	0.00 (−1.75-5.75)	148	11.0 (0.00-29.2)	3678	<.001	−0.366
Class 9	98	0.00 (−2.00-10.0)	148	14.0 (2.00-26.7)	3876	<.001	−0.395
Class 10	103	0.00 (−5.00-9.00)	148	9.00 (0.25-18.7)	4635	<.001	−0.335
Class 11	95	0.00 (−5.00-2.00)	148	9.50 (0.00-21.0)	3389	<.001	−0.439
Class 12	97	0.00 (−2.00-9.00)	148	9.50 (0.00-20.0)	4541	<.001	−0.312

[a]Data are presented as medians (25th-75th percentile).

[b]Mann-Whitney *U* test.

[c]Effect size correlation for Mann-Whitney *U* test.

Discussion

Principal Findings

The primary aim of this study was to evaluate the effectiveness of a mobile app promoting stress management and well-being for working women. To this end, we conducted a 2-arm randomized controlled pragmatic trial with female employees at a private hospital using a long period program (8 week), a self-selected sample, and standardized questionnaires. The pragmatic trial is designed to test the effectiveness of an intervention in everyday life to maximize applicability and generalizability [55]. Groups were homogeneous regarding baseline characteristics, and although both groups showed a significant increase in general well-being, only the intervention group presented a significant increase on the well-being index (WHO-5 scores) and work-related well-being, as well as significant decreases in perceived stress (PSS-10 scores), work-related stress, and general stress. In addition, participants who used the well-being mobile app reported a significantly

greater reduction in stress levels and a significant increase in well-being levels after each daily activity than participants who used the active control app. These results indicate that the well-being app was more effective at reducing employee stress and improving well-being levels.

One of our primary outcome measures was the PSS-10, the most widely used psychological instrument for measuring the perception of stress. Our results showed that participants from both groups were severely distressed at baseline scores (control [$\bar{x}$=22.9; SD 6.24] and intervention [$\bar{x}$=22.6; SD 7.16]), and only the intervention group presented a clinically significant change in the PSS-10 score (−6.15 points, on average) at the end of the 8-week study period, reaching symptom-free status (PSS-10=15.6), as defined by Ebert et al [48].

The other primary outcome measure used in this study was the WHO-5, one of the most widely used questionnaires to assess subjective well-being. The WHO-5 baseline score for both groups was approximately 43%, which indicates reduced well-being. After 8 weeks, the control group had a WHO-5 increase of approximately 6%, whereas improvement for the intervention group was almost 17%. Therefore, the difference in WHO-5 scores between groups was approximately 11%, which is clinically significant. However, at the end of the 8-week study period, the participants in the intervention group still had mean WHO-5 values ($\bar{x}$=14.93; 59.72%) below the general population reference level of 73.37% [45].

In the field of stress research, many studies use the WHO-5 to assess a wide variety of aspects including links between working conditions and well-being [56], the association between psychosocial working conditions and psychological well-being [57], as well as the association between workplace stress and well-being [58]. Our results corroborate the findings of Goa et al [58], who found that approximately 35% (n=977) of a total of 2796 employees presented low well-being scores (WHO-5 < 50) and that women, younger workers (<40 years), workers with high educational levels, and those with higher levels of job stress reported higher rates of poor mental health.

Several factors may contribute to the results obtained in this study. First, we applied an app-based EMI to intervene and interact with the participants in their work environment and during moments of their daily life. In addition, our system has been designed to ensure that participants respond to questions *in the moment*, that is, soon after being alerted to do so. This is a great advantage in relation to (1) surveys that require people to make retrospective and often generalized judgments, which tend to be affected by memory limitations and recall biases [59], and (2) laboratory experiments that do not occur within the context of a person's daily life, as we know that context can influence a person's states and responses [60]. Self-monitoring has long been known to raise self-awareness and promote positive behavioral development under certain conditions [61,62]. Specifically, it has been theorized that being asked questions about one's momentary states, experiences, behaviors, and/or thoughts close to the time and context of their occurrence may help one become more mindful of their occurrence, thereby providing opportunity for change [63]. Although it was only recorded as self-monitoring, participants in the control group reported that measuring stress and well-being levels before and after 20 min made them aware of changes in their physiological state (ie, they became more aware of their thought processes and more reflective). This could explain the increase in general well-being also observed in the active control group.

Second, our 8-week training program was built on the basis of 2 central approaches aimed at reducing stress and improving well-being. As described previously, meditation was one of those foundations. Studies involving smartphone delivery of mindfulness interventions focusing on workplace stress [64,65], well-being [41,66-68], and depression [69] showed comparable results to previous traditional interventions focused on the same outcome variables. The advantage of the smartphone interventions is that they are more rigorous because of instruction standardization across participants in the experimental group, the inclusion of active control materials participants expected to benefit from, and objective measures of adherence (provided within the app) rather than self-report. The other foundation was positive psychology principles. A recent meta-analysis revealed that 67% of positive psychology interventions are delivered in a self-help format, sometimes in conjunction with face-to-face instruction and support [34]. It is already known that a *shotgun* approach involving multiple and different positive exercises may be more effective than engaging in only 1 activity [70]. In our well-being mobile app, we used a combination of exercises and information to develop skills across different positive psychology domains, besides providing information about the benefits of increased well-being. From a public health perspective, self-help interventions can serve as cost-effective mental health promotion tools to reach large target groups that may not otherwise be reached [71].

Third, the theoretical basis of the intervention was confined to evidence-based components, consistent with the recommendations presented by Bakker et al [27] to create better and more rigorous mental health apps (MHapps). The chosen strategies were as follows: (1) focusing on nonclinical mental health, psychological well-being, and coping abilities, aiming to increase accessibility, enable preventive use, and reduce stigma, therefore avoiding the harmful effects of using mental illness labels [72]; (2) using self-monitoring and self-reflection—core features of many evidence-based psychological therapeutic techniques—to promote psychological growth and enable progress evaluations [73], with the advantage that MHapps make it possible for users to record self-monitoring data during their usual daily routines, while undergoing challenges or directly experiencing stressors [18]; (3) applying behavioral activation (ie, encourages individuals to engage in physiologically activating and psychologically rewarding activities) to boost self-efficacy, psychological well-being, and a repertoire of coping skills, since an app may promote self-discovery by encouraging an activity and then prompting reflection on the experience immediately after [74]; (4) presenting brief and passive psychoeducation to develop mental health literacy, in other words, to teach the participants about psychological processes underlying their distress and inform them about resources available to manage it [75]; (5) using real-time engagement to allow users to seek help for

psychological challenges at the time they experience distress or soon after, thus opening new learning opportunities and applying coping strategies in ecologically valid contexts [76]; (6) promoting activities explicitly linked to specific mood problems to enhance understanding of cause-and-effect relationships between actions and emotions [77]; (7) using gamification—the use of "game-based mechanics, aesthetics and game thinking to engage people, motivate action, promote learning and solve problems" [78]—and intrinsic motivation to encourage app use via rewards and internal triggers, positive reinforcement, and behavioral conditioning, emergent approaches that may help counteract motivation problems and yield additional well-being outcomes; (8) providing reminders (email and push notifications) as external triggers for engagement, aiming to increase adherence and reduce dropout from self-help interventions [79]; and (9) conducting an experimental trial to establish the app's efficacy before recommending it as an effective intervention.

Donker et al [26] systematically investigated the effectiveness of MHapps. Only 8 papers (describing 5 apps) were identified as providing scientific support for MHapps. Only 1 of these was a self-contained app, whereas the other 4 required input from a mental health professional. Unfortunately, none of these apps is currently available in app stores. These findings revealed the lack of experimental evidence for MHapps, of which hundreds are available. Despite the small number of studies included in their review, Donker et al [26] concluded that although apps have the potential to benefit those with poor mental health, further studies are needed to fully understand their usefulness and effectiveness. In summary, our results show that the intervention improved subjective symptoms of well-being and reduced stress through a wide array of evidence-based techniques that were delivered via a simple, enjoyable, intuitive, and interactive design. The simplicity of a program's interface and ease of navigation significantly influence users' perception of the quality of Web-based mental health interventions [80]. Our results corroborated the key outcomes observed by Coulon et al [81] that evidence-based, transparent, functional, and user-friendly apps may engage patients in effective stress reduction strategies.

Despite the proven efficacy of internet-based mobile-supported mental health interventions, high attrition rates are observed [66,82]. Nonadherence is a common issue in online psychological interventions and may reduce the effectiveness of an intervention [83]. A systematic review of attrition rates from internet-based intervention programs in which contact with a therapist was minimal reported an average dropout rate of 31% (range: 2%-83%) [84]. These data suggest that this trial had an acceptable attrition rate (54%) relative to other eHealth programs. Unfortunately, it is difficult to compare adherence levels of this trial with those of other internet-based stress management interventions because few studies thus far have reported this information. The percentage of participants in this study who completed both conditions (46%) was similar to available intervention completion rates (eg, 37%, [85]; 44%, [86]; and 88% [87]). The adherence rates of a meta-analysis of online MBIs varied between 35% and 92% [10].

Participants who adhered perceived the intervention as manageable and felt that they had learned useful strategies. On the other hand, many nonadherent participants declared that it was not easy to make time to do the exercises, although the self-help intervention helped them recognize the importance of self-regulating attention on their daily activities and finding personal space to relax. For this reason, it is critical to analyze the potential outcomes of traditional protocols delivered by new technologies. This critical investigation allows a better understanding of whether self-help mobile apps can be beneficial to users who are seeking well-being training to obtain better mental health.

Improving adherence is a high priority for internet interventions, as higher usage rates are associated with significant improvements in well-being [17]. Our mobile app used a *one size fits all* approach, which may not have been appropriate for a large group of people. Previous research indicates that providing support has a positive influence on adherence and enhances the effectiveness of online interventions [10,88]. Therefore, offering complementary instructor guidance to participants may potentially improve adherence and outcomes; however, instructor involvement is costly and may restrict the intervention's scalability. These barriers may be partially overcome by using automated support. Automated support has been shown to improve intervention adherence and effectiveness [89].

The authors would like to highlight that this is a well-being and stress reduction educational app and that people suffering from mental disorders or whose symptoms are difficult to manage should seek support from a mental health professional. We would also like to emphasize that this app offers an initial experience in the field of self-care with introductory classes in meditation, relaxation, attention, and positive psychology topics. Although it may be an ideal tool for those who do not have time to attend regular classes, it is not intended to substitute the guidance of experienced/certified instructors for people interested in deepening their practice.

Limitations

To interpret the results reported in this paper, some limitations should be taken into account. First, the main limitation of the study was that participants were not blinded, which may have increased bias. However, the designed study provided evidence of a pragmatic self-care intervention. Second, for feasibility reasons, the answers were self-reported. Third, we included only women and the volunteers self-selected into the trial, which limits the generalizability of the results. These limitations, however, should have minimal impact on the validity of the data, given the number of strengths of this study (eg, large sample size, use of an active control group, objective measures of intervention adherence, app developed for both Google's Android and Apple's iOS mobile operating systems).

Future Research

Future research should aim to replicate the results of this trial and investigate variables that may affect outcome and adherence. Personalization and tailoring intervention to individual needs, from baseline stress and well-being levels, and interactive

support could contribute to increased adherence and thereby to boost the effectiveness of the intervention. As all responses here were self-reported, integrating psychophysiological sensors with the apps and collecting biological measures, such as cortisol levels, would also strengthen the evidence of the beneficial effects of the well-being mobile app. Future studies are also necessary to investigate larger and more heterogeneous samples (in healthy subjects as well as clinical populations), to assess long-term efficacy measures (eg, including follow-ups of 1 or more years), and to better establish specific effects of this particular well-being mobile-based program.

Moreover, it would be interesting to test this protocol against the gold standard in the field (ie, face-to-face interventions) and assess which training format works best for which type of participant and under which circumstances. Although both formats may be equally effective, they may work differently on participants with different personal characteristics, and mobile-based interventions may be more advantageous in terms of efficiency and costs.

It is important to note that no single organization regulates parameters of app efficacy or transparency. The abundance of apps that have promising descriptions but are ultimately not well developed or maintained contributes to app overload and renders people vulnerable to misdirection and discouragement as they attempt to select and adopt effective self-management strategies. Well-planned studies will improve our ability to reach the potential of mobile mental health on both individual and population levels.

Conclusions

Our results indicate that the well-being and stress reduction app was better than the active control app at reducing employee stress and improving well-being levels. This trial contributes to the limited evidence available regarding the feasibility and efficacy of mobile-supported stress management and well-being interventions. Moreover, it is the first study to include an active control app group. The well-being mobile app presented here was highly effective in reducing perceived stress and improving mental well-being indices over a period of 4 and 8 weeks among women working at a private hospital. Thus, self-care mobile-based interventions may be used as preventive, easily accessible, and nonstigmatizing tools in a public health environment.

Acknowledgments

This research was supported with a grant received from Natura Cosméticos SA, Brazil, and Instituto Israelita de Ensino e Pesquisa Albert Einstein. The funders had no role in data collection or analyses, the decision to publish, or the preparation of the manuscript. CC received a doctoral fellowship from CAPES (Brazilian Federal Agency for the Improvement of Higher Education). EHK receives a research fellowship from CNPq (National Council of Scientific and Technological Development). The authors would like to acknowledge the volunteers who were willing to participate in this research and would like to thank their colleagues, friends, and family for their valuable support, guidance, and feedback throughout the project.

Authors' Contributions

All authors contributed to the study design, app design, and features. AL contributed to software development. CC and AL collected the data and organized the database. CC and SL conducted the statistical analyses. CC drafted the manuscript, and all authors approved the final version of the manuscript. EHK supervised the study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-EHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 2390 KB - jmir_v21i11e14269_app1.pdf]

Multimedia Appendix 2
Screenshots of the well-being mobile app, showing the app on Apple's App Store, user log-in page, pop-up messages, evaluations, home page, brief theoretical page, guided practice timer, and profile page. Image was produced by the first author.
[PNG File , 790 KB - jmir_v21i11e14269_app2.png]

Multimedia Appendix 3
Screenshots of the active control app, showing the app on Apple's App Store, user log-in page, pop-up messages, evaluations, home page, and pre- and post evaluation separated by a 20-min interval. Image was produced by first author.
[PNG File , 695 KB - jmir_v21i11e14269_app3.png]

Multimedia Appendix 4
Baseline outcome score data by experimental group.

[PDF File (Adobe PDF File), 197 KB - jmir_v21i11e14269_app4.pdf]

Multimedia Appendix 5
Preintervention × midintervention between-group comparisons.
[PDF File (Adobe PDF File), 287 KB - jmir_v21i11e14269_app5.pdf]

Multimedia Appendix 6
Preintervention × midintervention × postintervention between-group comparisons.
[PDF File (Adobe PDF File), 292 KB - jmir_v21i11e14269_app6.pdf]

References

1. Colligan TW, Higgins EM. Workplace stress: etiology and consequences. J Workplace Behav Health 2006 Jul 25;21(2):89-97. [doi: 10.1300/j490v21n02_07]
2. Wellens BT, Smith AP. Combined workplace stressors and their relationship with mood, physiology, and performance. Work Stress 2006 Jul;20(3):245-258. [doi: 10.1080/02678370601022712]
3. Milner A, King T, LaMontagne AD, Bentley R, Kavanagh A. Men's work, women's work, and mental health: a longitudinal investigation of the relationship between the gender composition of occupations and mental health. Soc Sci Med 2018 May;204:16-22. [doi: 10.1016/j.socscimed.2018.03.020] [Medline: 29550631]
4. American Psychological Association. 2017. Stress in America: The State of Our Nation URL: https://www.apa.org/images/state-nation_tcm7-225609.pdf [accessed 2019-02-21]
5. Burke RJ. Work stress and women's health: occupational status effects. J Bus Ethics 2002;37(1):91-102. [doi: 10.1023/A:1014734302972]
6. Olsen IB, Øverland S, Reme SE, Løvvik C. Exploring work-related causal attributions of common mental disorders. J Occup Rehabil 2015 Sep;25(3):493-505 [FREE Full text] [doi: 10.1007/s10926-014-9556-z] [Medline: 25451074]
7. Jesse R. Global Gender Gap Report. Switzerland: World Economic Forum; 2016.
8. Kabat-Zinn J. Mindfulness-based interventions in context: past, present, and future. Clin Psychol Sci Pr 2003;10(2):144-156. [doi: 10.1093/clipsy/bpg016]
9. Hölzel BK, Lazar SW, Gard T, Schuman-Olivier Z, Vago DR, Ott U. How does mindfulness meditation work? Proposing mechanisms of action from a conceptual and neural perspective. Perspect Psychol Sci 2011 Nov;6(6):537-559. [doi: 10.1177/1745691611419671] [Medline: 26168376]
10. Spijkerman MP, Pots WT, Bohlmeijer ET. Effectiveness of online mindfulness-based interventions in improving mental health: a review and meta-analysis of randomised controlled trials. Clin Psychol Rev 2016 Apr;45:102-114 [FREE Full text] [doi: 10.1016/j.cpr.2016.03.009] [Medline: 27111302]
11. Mind Your Mind Project. 2001. Mental Health: Strengthening Mental Health Promotion URL: https://mindyourmindproject.org/wp-content/uploads/2014/11/WHO-Statement-on-Mental-Health-Promotion.pdf [accessed 2019-09-14]
12. Seligman ME. Flourish: A Visionary New Understanding of Happiness and Well-Being. New York: Free Press; 2011.
13. Keyes CL, Dhingra SS, Simoes EJ. Change in level of positive mental health as a predictor of future risk of mental illness. Am J Public Health 2010 Dec;100(12):2366-2371. [doi: 10.2105/AJPH.2010.192245] [Medline: 20966364]
14. Wood AM, Joseph S. The absence of positive psychological (eudemonic) well-being as a risk factor for depression: a ten year cohort study. J Affect Disord 2010 May;122(3):213-217. [doi: 10.1016/j.jad.2009.06.032] [Medline: 19706357]
15. Sin NL, Lyubomirsky S. Enhancing well-being and alleviating depressive symptoms with positive psychology interventions: a practice-friendly meta-analysis. J Clin Psychol 2009 May;65(5):467-487. [doi: 10.1002/jclp.20593] [Medline: 19301241]
16. Bolier L, Haverman M, Kramer J, Westerhof GJ, Riper H, Walburg JA, et al. An internet-based intervention to promote mental fitness for mildly depressed adults: randomized controlled trial. J Med Internet Res 2013 Sep 16;15(9):e200 [FREE Full text] [doi: 10.2196/jmir.2603] [Medline: 24041479]
17. Manicavasagar V, Horswood D, Burckhardt R, Lum A, Hadzi-Pavlovic D, Parker G. Feasibility and effectiveness of a web-based positive psychology program for youth mental health: randomized controlled trial. J Med Internet Res 2014 Jun 4;16(6):e140 [FREE Full text] [doi: 10.2196/jmir.3176] [Medline: 24901900]
18. Carissoli C, Villani D, Riva G. Does a meditation protocol supported by a mobile application help people reduce stress? Suggestions from a controlled pragmatic trial. Cyberpsychol Behav Soc Netw 2015 Jan;18(1):46-53. [doi: 10.1089/cyber.2014.0062] [Medline: 25584730]
19. World Health Organization. Mental Health Action Plan 2013 - 2020. Geneva, Switzerland: World Health Organization; 2013.
20. Free C, Phillips G, Galli L, Watson L, Felix L, Edwards P, et al. The effectiveness of mobile-health technology-based health behaviour change or disease management interventions for health care consumers: a systematic review. PLoS Med 2013;10(1):e1001362 [FREE Full text] [doi: 10.1371/journal.pmed.1001362] [Medline: 23349621]
21. Kemp S. DataReportal – Global Digital Insights. 2019. Digital 2019: Q2 Global Digital Statshot URL: https://datareportal.com/reports/digital-2019-q2-global-digital-statshot [accessed 2019-05-31]

22. Kemp S. DataReportal – Global Digital Insights. 2019. Digital 2019: Brazil URL: https://datareportal.com/reports/digital-2019-brazil [accessed 2019-05-31]
23. IQVIA Institute for Human Data Science. 2017. The Growing Value of Digital Health: Evidence and Impact on Human Health and the Healthcare System URL: https://www.iqvia.com/institute/reports/the-growing-value-of-digital-health [accessed 2019-02-11]
24. IQVIA. 2015. Patient Adoption of mHealth: Use, Evidence and Remaining Barriers to Mainstream Acceptance URL: https://www.iqvia.com/-/media/iqvia/pdfs/institute-reports/patient-adoption-of-mhealth.pdf [accessed 2019-02-11]
25. Killingsworth MA, Gilbert DT. A wandering mind is an unhappy mind. Science 2010 Nov 12;330(6006):932 [FREE Full text] [doi: 10.1126/science.1192439] [Medline: 21071660]
26. Donker T, Petrie K, Proudfoot J, Clarke J, Birch M, Christensen H. Smartphones for smarter delivery of mental health programs: a systematic review. J Med Internet Res 2013 Nov 15;15(11):e247 [FREE Full text] [doi: 10.2196/jmir.2791] [Medline: 24240579]
27. Bakker D, Kazantzis N, Rickwood D, Rickard N. Mental health smartphone apps: review and evidence-based recommendations for future developments. JMIR Ment Health 2016 Mar 1;3(1):e7 [FREE Full text] [doi: 10.2196/mental.4984] [Medline: 26932350]
28. McKay FH, Cheng C, Wright A, Shill J, Stephens H, Uccellini M. Evaluating mobile phone applications for health behaviour change: a systematic review. J Telemed Telecare 2018 Jan;24(1):22-30. [doi: 10.1177/1357633X16673538] [Medline: 27760883]
29. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. Br Med J 2010 Mar 23;340:c332 [FREE Full text] [doi: 10.1136/bmj.c332] [Medline: 20332509]
30. Boutron I, Altman DG, Moher D, Schulz KF, Ravaud P, CONSORT NPT Group. CONSORT statement for randomized trials of nonpharmacologic treatments: a 2017 update and a CONSORT extension for nonpharmacologic trial abstracts. Ann Intern Med 2017 Jul 4;167(1):40-47. [doi: 10.7326/M17-0046] [Medline: 28630973]
31. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126. [doi: 10.2196/jmir.1923] [Medline: 22209829]
32. Dusek JA, Benson H. Mind-body medicine: a model of the comparative clinical impact of the acute stress and relaxation responses. Minn Med 2009 May;92(5):47-50 [FREE Full text] [Medline: 19552264]
33. Seligman ME, Csikszentmihalyi M. Positive psychology. An introduction. Am Psychol 2000 Jan;55(1):5-14. [doi: 10.1037//0003-066x.55.1.5] [Medline: 11392865]
34. Bolier L, Haverman M, Westerhof GJ, Riper H, Smit F, Bohlmeijer E. Positive psychology interventions: a meta-analysis of randomized controlled studies. BMC Public Health 2013 Feb 8;13:119 [FREE Full text] [doi: 10.1186/1471-2458-13-119] [Medline: 23390882]
35. Salvagioni DA, Melanda FN, Mesas AE, González AD, Gabani FL, Andrade SM. Physical, psychological and occupational consequences of job burnout: a systematic review of prospective studies. PLoS One 2017;12(10):e0185781 [FREE Full text] [doi: 10.1371/journal.pone.0185781] [Medline: 28977041]
36. Kabat-Zinn J. Full Catastrophe Living: Using the Wisdom of Your Body and Mind to Face Stress, Pain, and Illness. New York: Dell Publishing; 1990.
37. Kabat-Zinn J. Wherever You Go, There You Are: Mindfulness Meditation for Everyday Life. New York: Hyperion; 1994.
38. Wallace BA. Meditations of a Buddhist Skeptic: A Manifesto for the Mind Sciences and Contemplative Practice. New York, United States: Columbia University Press; 2012.
39. Chiesa A, Calati R, Serretti A. Does mindfulness training improve cognitive abilities? A systematic review of neuropsychological findings. Clin Psychol Rev 2011 Apr;31(3):449-464. [doi: 10.1016/j.cpr.2010.11.003] [Medline: 21183265]
40. Torous J, Firth J. The digital placebo effect: mobile mental health meets clinical psychiatry. Lancet Psychiatry 2016 Feb;3(2):100-102. [doi: 10.1016/S2215-0366(15)00565-9] [Medline: 26851322]
41. Howells A, Ivtzan I, Eiroa-Orosa FJ. Putting the 'app' in happiness: a randomised controlled trial of a smartphone-based mindfulness intervention to enhance wellbeing. J Happiness Stud 2014 Oct 29;17(1):163-185. [doi: 10.1007/s10902-014-9589-1]
42. Cohen S, Williamson G. Perceived stress in a probability sample of the United States. In: Spacapan S, Oskamp S, editors. The Social Psychology of Health: The Claremont Symposium on Applied Social Psychology. Thousand Oaks, CA: Sage Publications; 1988:31-67.
43. Reis RS, Hino AA, Añez CR. Perceived stress scale: reliability and validity study in Brazil. J Health Psychol 2010 Jan;15(1):107-114. [doi: 10.1177/1359105309346343] [Medline: 20064889]
44. Staehr JK. WHO/Europe. 1998. Wellbeing Measures in Primary Health Care/ The Depcare Project: Report on a WHO Meeting URL: http://www.euro.who.int/__data/assets/pdf_file/0016/130750/E60246.pdf [accessed 2019-09-14]
45. de Souza CM, Hidalgo MP. World Health Organization 5-item well-being index: validation of the Brazilian Portuguese version. Eur Arch Psychiatry Clin Neurosci 2012 Apr;262(3):239-244. [doi: 10.1007/s00406-011-0255-x] [Medline: 21912931]

46. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav 1983 Dec;24(4):385-396. [doi: 10.2307/2136404] [Medline: 6668417]
47. Heber E, Lehr D, Ebert DD, Berking M, Riper H. Web-based and mobile stress management intervention for employees: a randomized controlled trial. J Med Internet Res 2016 Jan 27;18(1):e21 [FREE Full text] [doi: 10.2196/jmir.5112] [Medline: 26818683]
48. Ebert DD, Kählke F, Buntrock C, Berking M, Smit F, Heber E, et al. A health economic outcome evaluation of an internet-based mobile-supported stress management intervention for employees. Scand J Work Environ Health 2018 Mar 1;44(2):171-182. [doi: 10.5271/sjweh.3691] [Medline: 29144535]
49. Lesage F, Berjot S, Deschamps F. Psychometric properties of the French versions of the perceived stress scale. Int J Occup Med Environ Health 2012 Jun;25(2):178-184 [FREE Full text] [doi: 10.2478/S13382-012-0024-8] [Medline: 22528542]
50. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. J Consult Clin Psychol 1991 Feb;59(1):12-19. [doi: 10.1037//0022-006x.59.1.12] [Medline: 2002127]
51. Löwe B, Spitzer RL, Gräfe K, Kroenke K, Quenter A, Zipfel S, et al. Comparative validity of three screening questionnaires for DSM-IV depressive disorders and physicians' diagnoses. J Affect Disord 2004 Feb;78(2):131-140. [doi: 10.1016/s0165-0327(02)00237-9] [Medline: 14706723]
52. Topp CW, Østergaard SD, Søndergaard S, Bech P. The WHO-5 Well-Being Index: a systematic review of the literature. Psychother Psychosom 2015;84(3):167-176 [FREE Full text] [doi: 10.1159/000376585] [Medline: 25831962]
53. Florescer - Bem Vindo ao Programa. URL: http://florescer.inovaeinstein.com.br/Site/BemVindo.php [accessed 2019-02-25]
54. Amazon Web Services. Perguntas Frequentes Sobre O Aws Elastic Beanstalk URL: https://aws.amazon.com/pt/elasticbeanstalk/faqs/ [accessed 2019-02-25]
55. Patsopoulos NA. A pragmatic view on pragmatic trials. Dialogues Clin Neurosci 2011;13(2):217-224 [FREE Full text] [Medline: 21842619]
56. Nielsen K, Randall R. The importance of employee participation and perceptions of changes in procedures in a teamworking intervention. Work Stress 2012 Apr;26(2):91-111 [FREE Full text] [doi: 10.1080/02678373.2012.682721] [Medline: 22745519]
57. Schütte S, Chastang JF, Malard L, Parent-Thirion A, Vermeylen G, Niedhammer I. Psychosocial working conditions and psychological well-being among employees in 34 European countries. Int Arch Occup Environ Health 2014 Nov;87(8):897-907. [doi: 10.1007/s00420-014-0930-0] [Medline: 24526111]
58. Gao J, Weaver SR, Dai J, Jia Y, Liu X, Jin K, et al. Workplace social capital and mental health among Chinese employees: a multi-level, cross-sectional study. PLoS One 2014;9(1):e85005 [FREE Full text] [doi: 10.1371/journal.pone.0085005] [Medline: 24404199]
59. Schwarz N. Retrospective and concurrent self-reports: the rationale for real-time data capture. In: Stone A, Shiffman S, Atienza A, Nebeling L, editors. The Science of Real-Time Data Capture: Self-Reports in Health Research. Oxford, United Kingdom: Oxford University Press; 2007:471.
60. Wilhelm P, Perrez M, Pawlik K. Conducting research in daily life: a historical review. In: Mehl MR, Conner TS, Csikszentmihalyi M, editors. Handbook of Research Methods for Studying Daily Life. New York: Guilford Press; 2011:108-123.
61. Korotitsch WJ, Nelson-Gray RO. An overview of self-monitoring research in assessment and treatment. Psychol Assess 1999;11(4):415-425. [doi: 10.1037/1040-3590.11.4.415]
62. Cohen JS, Edmunds JM, Brodman DM, Benjamin CL, Kendall PC. Using self-monitoring: implementation of collaborative empiricism in cognitive-behavioral therapy. Cogn Behav Pract 2013 Nov;20(4):419-428. [doi: 10.1016/j.cbpra.2012.06.002]
63. Goodwin MS, Velicer WF, Intille SS. Telemetric monitoring in the behavior sciences. Behav Res Methods 2008 Feb;40(1):328-341. [doi: 10.3758/BRM.40.1.328] [Medline: 18411557]
64. Bostock SK, Steptoe A. Can Finding Headspace Reduce Work Stress? A Randomised Controlled Workplace Trial of a Mindfulness Meditation App. In: Proceedings of the 71st Annual Scientific Meeting of the American-Psychosomatic-Society. 2013 Presented at: APS'13; March 13-16, 2013; Miami, Florida p. A36-A37.
65. Bostock S, Crosswell AD, Prather AA, Steptoe A. Mindfulness on-the-go: effects of a mindfulness meditation app on work stress and well-being. J Occup Health Psychol 2019 Feb;24(1):127-138. [doi: 10.1037/ocp0000118] [Medline: 29723001]
66. Mak WW, Tong AC, Yip SY, Lui WW, Chio FH, Chan AT, et al. Efficacy and moderation of mobile app-based programs for mindfulness-based training, self-compassion training, and cognitive behavioral psychoeducation on mental health: randomized controlled noninferiority trial. JMIR Ment Health 2018 Oct 11;5(4):e60 [FREE Full text] [doi: 10.2196/mental.8597] [Medline: 30309837]
67. van Emmerik AA, Berings F, Lancee J. Efficacy of a mindfulness-based mobile application: a randomized waiting-list controlled trial. Mindfulness (N Y) 2018;9(1):187-198 [FREE Full text] [doi: 10.1007/s12671-017-0761-7] [Medline: 29387266]
68. Walsh KM, Saab BJ, Farb NA. Effects of a mindfulness meditation app on subjective well-being: active randomized controlled trial and experience sampling study. JMIR Ment Health 2019 Jan 8;6(1):e10844 [FREE Full text] [doi: 10.2196/10844] [Medline: 30622094]

69. Ly KH, Trüschel A, Jarl L, Magnusson S, Windahl T, Johansson R, et al. Behavioural activation versus mindfulness-based guided self-help treatment administered through a smartphone application: a randomised controlled trial. BMJ Open 2014 Jan 9;4(1):e003440 [FREE Full text] [doi: 10.1136/bmjopen-2013-003440] [Medline: 24413342]
70. Seligman ME, Steen TA, Park N, Peterson C. Positive psychology progress: empirical validation of interventions. Am Psychol 2005;60(5):410-421. [doi: 10.1037/0003-066X.60.5.410] [Medline: 16045394]
71. den Boer PC, Wiersma D, van den Bosch RJ. Why is self-help neglected in the treatment of emotional disorders? A meta-analysis. Psychol Med 2004 Aug;34(6):959-971. [doi: 10.1017/s003329170300179x] [Medline: 15554567]
72. Corrigan PW, Watson AC, Barr L. The self–stigma of mental illness: implications for self–esteem and self–efficacy. J Clin Psychol 2006 Oct;25(8):875-884. [doi: 10.1521/jscp.2006.25.8.875]
73. Humphreys KL, Marx BP, Lexington JM. Self-monitoring as a treatment vehicle. In: O'Donohue WT, Fisher JE, editors. Cognitive Behavior Therapy: Core Principles for Practice. Ssecond Edition. Hoboken, NJ: John Wiley & Sons Inc; 2009:576-583.
74. Martell CR, Dimidjian S, Lewinsohn PM. Behavioral activation therapy. In: Kazantzis N, Reinecke MA, Freeman A, editors. Cognitive and Behavioral Theories in Clinical Practice. New York: Guilford Press; 2010:198-223.
75. Macrodimitris SD, Hamilton KE, Backs-Dermott BJ, Mothersill KJ. CBT basics: a group approach to teaching fundamental cognitive-behavioral skills. J Cogn Psychother 2010 May 1;24(2):132-146. [doi: 10.1891/0889-8391.24.2.132]
76. Heron KE, Smyth JM. Ecological momentary interventions: incorporating mobile technology into psychosocial and health behaviour treatments. Br J Health Psychol 2010 Feb;15(Pt 1):1-39 [FREE Full text] [doi: 10.1348/135910709X466063] [Medline: 19646331]
77. Wendel S. Designing for Behavior Change: Applying Psychology and Behavioral Economics. Sebastopol, CA: O'Reilly Media Inc; 2013.
78. Kapp KM. The Gamification of Learning and Instruction: Game–Based Methods and Strategies for Training and Education. First Edition. San Francisco, CA: Pfeiffer; 2012.
79. Cavanagh K. Turn on, tune in and (don't) drop out: engagement, adherence, attrition, and alliance with internet-based interventions. In: Bennett-Levy J, Richards D, Farrand P, Christensen H, Griffiths K, Kavanagh D, et al, editors. Oxford Guide to Low Intensity CBT Interventions. First Edition. Oxford, UK: Oxford University Press; 2010:227-233.
80. Kerr C, Murray E, Stevenson F, Gore C, Nazareth I. Internet interventions for long-term conditions: patient and caregiver quality criteria. J Med Internet Res 2006 Jul 28;8(3):e13 [FREE Full text] [doi: 10.2196/jmir.8.3.e13] [Medline: 16954123]
81. Coulon SM, Monroe CM, West DS. A systematic, multi-domain review of mobile smartphone apps for evidence-based stress management. Am J Prev Med 2016 Jul;51(1):95-105. [doi: 10.1016/j.amepre.2016.01.026] [Medline: 26993534]
82. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
83. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. J Med Internet Res 2011 Aug 5;13(3):e52 [FREE Full text] [doi: 10.2196/jmir.1772] [Medline: 21821503]
84. Melville KM, Casey LM, Kavanagh DJ. Dropout from internet-based treatment for psychological disorders. Br J Clin Psychol 2010 Nov;49(Pt 4):455-471. [doi: 10.1348/014466509X472138] [Medline: 19799804]
85. Drozd F, Raeder S, Kraft P, Bjørkli CA. Multilevel growth curve analyses of treatment effects of a web-based intervention for stress reduction: randomized controlled trial. J Med Internet Res 2013 Apr 22;15(4):e84 [FREE Full text] [doi: 10.2196/jmir.2570] [Medline: 23607962]
86. Ly KH, Asplund K, Andersson G. Stress management for middle managers via an acceptance and commitment-based smartphone application: a randomized controlled trial. Internet Interv 2014 Jul;1(3):95-101. [doi: 10.1016/j.invent.2014.06.003]
87. Rose RD, Buckey JC, Zbozinek TD, Motivala SJ, Glenn DE, Cartreine JA, et al. A randomized controlled trial of a self-guided, multimedia, stress management and resilience training program. Behav Res Ther 2013 Feb;51(2):106-112. [doi: 10.1016/j.brat.2012.11.003] [Medline: 23262118]
88. Brouwer W, Kroeze W, Crutzen R, de Nooijer J, de Vries NK, Brug J, et al. Which intervention characteristics are related to more exposure to internet-delivered healthy lifestyle promotion interventions? A systematic review. J Med Internet Res 2011 Jan 6;13(1):e2 [FREE Full text] [doi: 10.2196/jmir.1639] [Medline: 21212045]
89. Furmark T, Carlbring P, Hedman E, Sonnenstein A, Clevberger P, Bohman B, et al. Guided and unguided self-help for social anxiety disorder: randomised controlled trial. Br J Psychiatry 2009 Nov;195(5):440-447. [doi: 10.1192/bjp.bp.108.060996] [Medline: 19880935]

Abbreviations

ΔSL: stress level variation
ΔWBL: well-being level variation
ANOVA: analysis of variance
AWS: Amazon Web Services

CI: confidence interval
CONSORT: Consolidated Standards of Reporting Trials
eHealth: electronic health
EMI: ecological momentary intervention
MBIs: mindfulness-based interventions
MHapps: mental health apps
mHealth: mobile health
PSS: perceived stress scale
PSS-10: 10-item perceived stress scale
WHO: World Health Organization
WHO-5: 5-item World Health Organization Well-Being Index
WMW: Wilcoxon-Mann-Whitney test

Edited by C Eichenberg; submitted 05.04.19; peer-reviewed by S Badawy, D Vilani, S Baptista, M Hart; comments to author 25.05.19; revised version received 25.06.19; accepted 01.08.19; published 07.11.19.

Please cite as:
Coelhoso CC, Tobo PR, Lacerda SS, Lima AH, Barrichello CRC, Amaro Jr E, Kozasa EH
A New Mental Health Mobile App for Well-Being and Stress Reduction in Working Women: Randomized Controlled Trial
J Med Internet Res 2019;21(11):e14269
URL: https://www.jmir.org/2019/11/e14269
doi:10.2196/14269
PMID:31697244

Review

Mobile and Web-Based Apps That Support Self-Management and Transition in Young People With Chronic Illness: Systematic Review

Yisselle Ilene Virella Pérez[1,2], BS, MD; Sharon Medlow[1,2], BA (Hons), PhD; Jane Ho[3], BMed, FRACP, MPH, DCH, GradCert (Allergic Diseases); Katharine Steinbeck[1,2], MBBS, PhD, FRACP

[1]The Children's Hospital at Westmead, Academic Department of Adolescent Medicine, Westmead, New South Wales, Australia

[2]The University of Sydney, Faculty of Medicine and Health, Sydney Medical School, Discipline of Child and Adolescent Health, Westmead, New South Wales, Australia

[3]Sydney Children's Hospital Network, Sydney Children's Hospital, Randwick, Centre for Adolescent and Young Adult Health, Randwick, New South Wales, Australia

Corresponding Author:
Yisselle Ilene Virella Pérez, BS, MD
The Children's Hospital at Westmead
Academic Department of Adolescent Medicine
Corner Hawkesbury Road and Hainsworth Street
Locked Bag 4001
Westmead, New South Wales, 2145
Australia
Phone: 61 2 9845 2507
Email: yisselle.virella@gmail.com

Abstract

Background: More adolescents with chronic physical illness are living into adulthood, and they require the development of proficient self-management skills to maintain optimal physical health as they transition into adult care services. It is often during this vulnerable transition period that deterioration in illness control is seen as a result of inadequate self-management skills and understanding of their chronic illness. Mobile technology has been proposed as an innovative opportunity to assist in improving the management of chronic conditions as young people transition to adult care services. Over the past 5 years, there has been a significant increase in research into the use of health-related apps.

Objective: This study aimed to evaluate the utility and effectiveness of mobile and Web-based health apps that support self-management and transition in young people with chronic physical health illnesses.

Methods: We conducted a comprehensive review of the literature in 5 bibliographic databases, using key search terms, considering only articles published from 2013, as we were extending the data from 2 previous systematic reviews. Abstracts were screened for possible inclusion by 2 reviewers. Data extraction and quality assessment tools were used for the evaluation of included studies.

Results: A total of 1737 records were identified from the combined electronic searches, and 854 records were removed as duplicates. A total of 68 full articles were further assessed for eligibility, and 6 articles met our review criteria: 3 pilot studies, 2 randomized controlled trials, and 1 prospective cohort study. Publication years ranged from 2015 to 2018. The apps reported were targeted at type 1 diabetes mellitus, epilepsy, asthma, beta thalassemia major, and sickle cell disease, with a combined sample size of 336. A total of 4 studies included in this review reported being effective in increasing knowledge of the targeted condition and increasing therapy adherence, including increased medication adherence. A total of 2 manuscripts only mentioned the word transition. Participant's satisfaction was reported for all studies. Heterogeneity of the studies prevented meta-analysis.

Conclusions: There remain limited data on the effectiveness and use of mobile and Web-based apps, which might facilitate the transition of adolescents with chronic illnesses from pediatric to adult health care services. This systematic review provides an updated overview of available apps for adolescents with chronic illnesses. This systematic review has been unable to provide evidence for effectiveness of this approach, but it does provide insights into future study design, with reference to the development, evaluation, and efficacy of apps tailored for adolescents with chronic illnesses, including the involvement of adolescents in such designs.

Trial Registration: PROSPERO CRD42018104611; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=104611

(J Med Internet Res 2019;21(11):e13579) doi:10.2196/13579

KEYWORDS

adolescent; mobile app; Web-based app; chronic illness; self-management; transition to adult care

Introduction

With advances in medical science, more adolescents with chronic physical illness are now living into adulthood [1-3]. Chronic illnesses, by definition, are long-term health conditions that, if incorrectly managed or left untreated, are likely to have consequences for the overall well-being of the person [4]. Chronic illnesses have a substantial impact on health care systems, as patients are frequent users of health care resources, and they often require complex interventions and treatment to manage their condition over a lifetime [5]. The development of self-management skills is essential for adolescents living with chronic conditions to maintain optimal physical health as they take responsibility for their own health care. Barlow et al define self-management as "the individual's ability to manage the symptoms, treatment, physical and psychosocial consequences and lifestyle changes inherent in living with a chronic condition" [6]. Self-management skills are of particular importance for adolescents with a chronic illness as they transition into adult clinical care, with parents and caregivers no longer taking responsibility for these tasks and adult health care professionals expecting young people to come equipped with these skills [7].

Transition is defined as the planned movement of adolescents with chronic illness from pediatric-centered care to adult health care systems [8]. It is during this time that deterioration in illness control and even unplanned hospitalizations are often observed as a result of inadequate self-management and physical deterioration [9-12]. Therefore, it is important to find developmentally appropriate and accessible ways to encourage and promote the improvement of young people's self-management behaviors [9].

Smartphones and tablets are widely used by adolescents and young adults. According to the Pew Research Center Internet and Technology (2018), an estimated 95% of adolescents in the United States own or have access to a smartphone, representing a 22% increase from 2014 to 2015. A total of 45% adolescents reported that they use the internet constantly on their device, nearly doubling the number from the 2014 to 2015 survey [13]. In the United Kingdom, 62% of adolescents aged 12 to 15 years own a mobile phone [14]. Similarly, 94% of Australians aged 16 to 17 years (age when active planning for the event of transfer should occur) owned a smartphone, and 89% of those between 18 and 24 years of age had a smartphone from which 83% of those downloaded an app [15], with numbers expected to continue rising. The multifunctional characteristics of smartphones and tablets allow multiple interventions and goals to be addressed, including knowledge enhancement and independent self-management skills [16]. It is not surprising that mobile and Web-based health technology has been proposed as a way to interact with young people with chronic illness as they enter the vulnerable transition period [8,17].

Currently, there are over 97,000 health-related apps available in the category of health and fitness in Google Play and the App Store, with approximately 1000 apps being created every month. It is estimated that the number of available apps will increase by about 25% each year [18]. Over the past 5 years, there has also been a significant increase in research on the utility of health-related apps [18].

To date, there has been little evidence produced about the use of mobile and Web-based apps aimed at supporting self-management in chronic illness and transition to adult care. A systematic review conducted by Majeed-Ariss et al [19] examined the literature published between 2003 and 2014 on the effectiveness of mobile apps designed to support adolescents' management of their physical chronic or long-term conditions. A total of 4 studies were included in the systematic review. Cafazzo et al's [20] and Frøisland et al's [21] apps were targeted at type 1 diabetes. Cafazzo et al's [20] pilot study trial showed that daily average frequency of blood glucose measurements increased by 50% (P=.006); however, glycated hemoglobin (HbA_{1c}) did not change significantly (P=.11). On the contrary, Frøisland et al's [21] study did not find any statistical significance in their outcomes, which were changes in HbA_{1c}, system usability, and theoretical knowledge. Burbank et al developed an app to improve asthma management [22]. Participant's satisfaction with the app was high (93%); asthma control tests scores improved significantly (P=.03), as well as asthma attack prevention self-efficacy scores (P=.04). Aldiss et al's app was targeted at cancer [23]. Even though they included 6 psychometric measures, the authors did not identify a primary outcome measure. The authors of the systematic review were unable to draw any conclusions, primarily because of the fact that the studies included in the review were generally in early proof-of-concept phase and with few participants.

A recent meta-analysis drawing on data published between 2006 and 2016 examined the use of mobile health interventions in improving health outcomes in young people [24]. This study showed that digital technology can be effective in eliciting meaningful improvements in pediatric health behavior and associated health outcomes. However, the average age of participants was only 11.4 years, and many of the included studies focused on interventions targeting care givers, thereby limiting conclusions around adolescents with chronic illnesses. Some of the health behaviors addressed included immunization adherence, HIV prevention, dental hygiene, sun safety, physical activity, smoking, obesity, diabetes, stem cell transplant, and asthma.

In summary, the usefulness or effectiveness of mobile health technology interventions among adolescents remains unclear. Adolescents are more likely to use the newest and smartest technology, which makes the use of mobile apps more

challenging. Technology evolves rapidly, and there is then the cost of competing with commercial developers who may be less interested in more niche markets [25], such as young people with chronic conditions.

The aims of this systematic review were the following: first, to evaluate the current available literature on the utility and effectiveness of mobile or Web-based health apps created for adolescents with chronic physical health conditions, which demonstrate some degree of user interaction over and above self-monitoring of the physical condition. Second, to describe features of the app that might explicitly or implicitly facilitate the transition of adolescents with chronic physical illness from pediatric to adult care.

Methods

This systematic review was registered with the International Prospective Register of Systematic Reviews (PROSPERO): CRD42018104611.

Search Strategy

A comprehensive literature search of 5 bibliographical databases was conducted to identify eligible studies. Databases included the following: Medical Literature Analysis and Retrieval System Online (MEDLINE), EMBASE, PsycINFO, Web of Science, and the Cumulative Index to Nursing and Allied Health Literature. The comprehensive search strategy was initially created in MEDLINE (Multimedia Appendix 1), and then it was adapted to the other databases. The specified chronic conditions were included to ensure that all available literature was included, and because these conditions require continuity of care and specialist-to-specialist transition. No restriction was placed on language. Only articles published from 2013 were considered, as this review is an update on the 2 previous systematic reviews cited in the Introduction section [19,24]. We also hypothesized that as general health-related app use research has increased in the past 5 years [18], this might also be reflected in an increase in adolescent studies. The end date for article searching was August 2018. Reference lists of relevant articles were hand searched to identify additional studies.

Inclusion and Exclusion Criteria

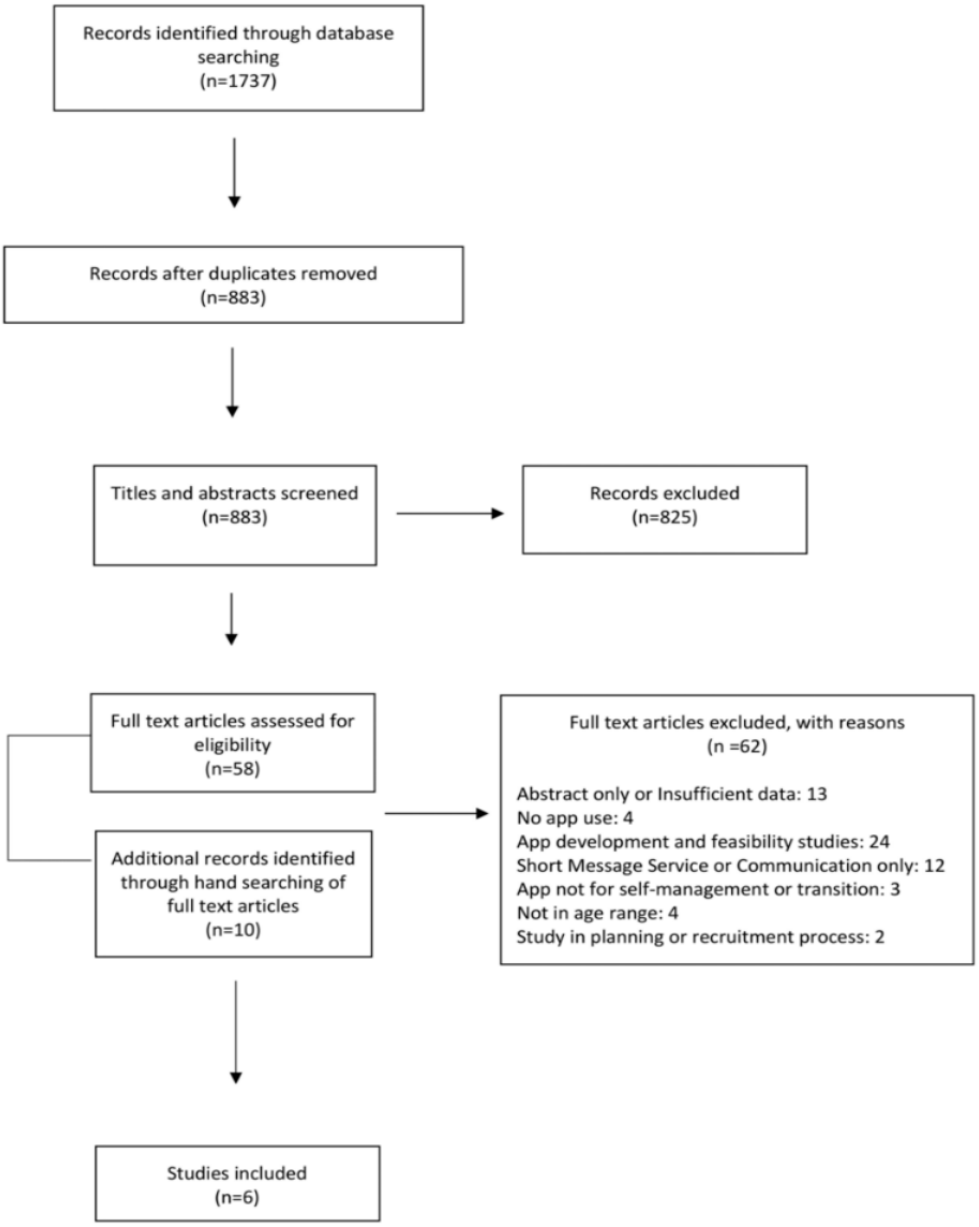
Only original quantitative and qualitative studies published in peer-reviewed journals were included. Studies were required to focus on the use of a mobile device or Web-based app as an intervention to support self-management and to aid transition of adolescents diagnosed with a chronic physical illness from pediatric to adult clinical care. Outcomes included any changes in behavioral, physiological, attitudinal, or knowledge variables. Adolescents were young people aged 10 to 19 years, as defined by the World Health Organization [26].

Studies that included mobile or Web-based apps for the management of mental health, acute cancer, pain, or lifestyle/health risk behaviors were excluded. Chronic pain syndromes and mental health are undeniably important health issues for adolescents, but the management approaches have some significant differences to those of chronic physical illness. Management in chronic physical illness is around long-term and often complex regimens, which highlight differences to their peers and where adequate control is often difficult to ascertain on a day-to-day basis. The strategies are likely to be different in type or dose to those employed in mental health and chronic pain, where rehabilitation and reframing of perceptions and thought are key, making it difficult to extrapolate this information to chronic physical illness. Studies were also excluded if the app was used only for monitoring health status, such as continuous glucose monitoring for diabetes or if less than 50% of the sample size of the study was in the adolescent age range of 10 to 19 years. It is acknowledged that the timing of transition varies among countries, but a majority of adolescents with chronic illness will have commenced the transfer to adult health services by the time that secondary schooling ends.

Study Selection

The design of this review followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses' (PRISMA) statement [27], and the PRISMA diagram for study selection is shown in Figure 1. Duplicate studies were removed. A total of 1 reviewer (YVP) performed the first-stage screening of titles and abstracts on the basis of the research question, as well as the inclusion and exclusion criteria. Abstracts were then screened for possible inclusion by 2 reviewers (YVP and KS) to determine full-text screening. There were no disagreements between these 2 reviewers, but JH had been identified as the third party to resolve consensus issues. Scientific abstracts were considered if these were published in a peer-reviewed journal and contained adequate data to apply inclusion and exclusion criteria. A total of 2 corresponding authors were contacted via email to obtain further information to be included in the final review of manuscripts. As no responses were received, these 2 abstracts were excluded.

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow diagram of the review.



Data Extraction

The information recorded included the following: demographics (age, gender, and country of origin), chronic physical illness, study characteristics (setting, study design, duration, data collection methodology, and sample size), and study outcomes. Any mention of transition was recorded. In addition, and when possible, further details of the app used to deliver an intervention were recorded: app name, platform, app's purpose, content, education delivered about its use, availability (eg, was an access code required), as well as parent and health care professional access to the app.

Quality Assessment

The checklist created by Downs and Black [28] for randomized and nonrandomized studies was used to assess the quality of the full manuscripts included in the systematic review. This scale is a 27-item checklist that assesses study reporting, external validity, internal validity in terms of bias and confounding (selection bias), and power. It has good reliability and high internal consistency [28]. The maximum score for the modified checklist was 28. Items were given 1 if the answer was "yes" or 0 if it was "no" or "unable to determine". A total of 1 item (Are the distributions of principal confounders in each group of subjects to be compared clearly described?) had a maximum possible score of 2. The score of item 27 was modified, as has been done in previous studies [29,30], awarding 1 or 0 points depending on whether there was statistical power to detect a clinically important effect. Score ranges were grouped into 4 categories: excellent (26-28), good (20-25), fair (15-19), and poor (≤14), as previously defined by Samoocha et al [30].

Statistical Analysis

Descriptive statistics were performed to capture the country where research was conducted, research setting, sample sizes, participant characteristics, health condition, study design, outcome measures, features, and usability of the apps. It was not possible to perform a meta-analysis because of the small

number of manuscripts that met inclusion criteria and the heterogeneity of study outcomes.

Results

The process of manuscript identification and selection is summarized in Figure 1. A total of 1737 records were identified from the combined electronic searches, from which 854 records were removed as duplicates. A total of 58 full-text articles, in addition to 10 articles identified through hand searching of full-text records' bibliographies, were selected for full-text review, and 6 articles met inclusion criteria for data extraction.

Characteristics of Included Studies

Characteristics of the 6 full manuscripts appear in Table 1.

Publication year ranged from 2015 to 2018. The studies describe 6 different apps used as interventions to support self-management or aid in transition of adolescents with different chronic physical illnesses: asthma [31,32], beta thalassemia major and sickle cell disease [33], type 1 diabetes mellitus [25,34], and epilepsy [16]. Of these manuscripts, 2 were pilot studies with an intervention group only [31,33], 1 was a pilot study with a control and intervention group [32], 2 were randomized controlled trials [25,34], and 1 was a prospective cohort study [16]. A total of 3 of the studies [25,32,34] compared 2 study groups, an intervention group in which the participants received usual care plus the intervention (app) and a control group that received only usual care; in 1 of these studies [32], the control group received a "control" version of the app to facilitate data transfer from an inhaler sensor. App features are summarized in Table 2. A total of 4 studies were conducted in North America, 1 in Australia and 1 in Denmark. All studies recruited participants from specialist clinics or pediatric hospitals. The setting details for the studies are outlined in Table 3. Sample sizes ranged from 7 to 151, with a combined sample size of 336. Participant ages ranged from 8 to 22 years, with a median of 14.1. The duration of the studies ranged from 1 to 12 months. There was an average 93% retention across studies.

Table 1. Characteristics of the included studies. The primary outcome measures were italicized for emphasis.

Author, Year	Country	Chronic condition	Initial/final	Duration (months)	Female/Male	Age (years), mean (range)	Study design	Outcome measures (*primary*)
Farooqui et al, 2015 [31]	United States	Asthma	24/21	1	9/12	11.6 (9-16)	Pilot Study	*MA ipod e-log*[a], *PI*[b]*Survey*
Cushing et al, 2016 [32]	United States	Asthma	7/5	3	5/2	14.1 (11-18)	Pilot Study	*US Interview* or *FG*[c], *RTMD*[d]
Goyal et al, 2017 [25]	Canada	Type 1 diabetes mellitus	92/91	12	51/41	14.1 (11-16)	Randomized controlled trial (RCT)	*HbA$_{1c}$*, SMBG[e], DQOLY[f], DFRQ[g], SCI[h], RTCS[i], 7-point LS [j], SS[k] interview
Leonard et al, 2017 [33]	United States	βThal[l] and SCD[m]	11/10	6	7/4	12.4 (8-21)	Pilot Study	*MA*[n]*self-videos*, *FQ*[o]; *KA*[p]
Castensøe -Seidenfaden et al, 2018 [34]	Denmark	Type 1 diabetes mellitus	151/148	12	81/70	17.6 (14-22)	RCT	*HbA$_{1c}$*, PCD[q], HCCQ[r], PAID[s]
Le Marne et al, 2018 [16]	Australia	Epilepsy	51/36	2[t]	27/24	14.5 (13-19)	Prospective Cohort	*SKEQ*[u], *AKEQ*[v], SSES-C[w], CATIS[x], MA[y]:parent log, MARS[z]

[a]MA ipod e-log: medication adherence assessed through electronic logging in iPod Touch.

[b]PI Survey: postintervention survey, in house, no details available.

[c]US Interview or FG: Unstructured interview or focus group to discuss experience with inhaler and app, no details available.

[d]RTMD: Real Time Medication Data from the sensor.

[e]SMBG: Self-monitoring Blood Glucose.

[f]DQOLY: Diabetes Quality of Life for Youth Questionnaire, to measure quality of life.

[g]DFRQ: Diabetes Family Responsibility Questionnaire, to measure adolescent-guardian interaction around care.

[h]SCI: Self-Care Inventory, to measure adherence to treatment recommendations.

[i]RTCS: Readiness to Change Survey, to assess participant self-management.

[j]7-point LS: 7-point Likert Scale, to assess overall satisfaction with the "bant" app.

[k]SS interview: semistructured interview, in house, no details available.

[l]βThal: beta thalassemia major.

[m]SCD: sickle cell disease.

[n]MA self-reported videos: medication adherence self-reported videos "selfies."

[o]FQ: Feasibility Questionnaire, to assess parents and participants feasibility, adherence, satisfaction, and ease of use of the Intensive Training Program mobile app.

[p]KA: Knowledge Assessment, to assess understanding of educational material presented in the modules.

[q]PCD: Perceived Competence in Diabetes, assesses patients' experience of being able to manage diabetes successfully.

[r]HCCQ: Health Care Climate Questionnaire, assesses the degree to which patients perceived their health care providers as supporting their autonomy.

[s]PAID: Problem Areas in Diabetes, assesses diabetes-related distress.

[t]2: mean duration 70 days (SD 39.9).

[u]SKEQ: Self Knowledge of Epilepsy Questionnaire modified version.

[v]AKEQ: Adolescent Knowledge of Epilepsy Questionnaire, assesses general epilepsy knowledge.

[w]SSES-C: Seizure Self-Efficacy Scale for Children and Adolescents with Epilepsy, measures self-reported efficacy in managing patient's own seizure disorder.

[x]CATIS: Child Attitude Toward Illness Scale, measures attitudes and feelings about having epilepsy.

[y]MA parent log: medication adherence logged by the parent or caregiver.

[z]MARS: Mobile Application Rating Scale, survey to collect feedback about the app.

Table 2. Features and usability of the apps.

Author, year	App name	App training	App content	Platform	App use
Farooqui et al, 2015 [31]	AsthmaCare	NR[a]	Daily reminders (medication use and personalized triggered avoidance strategies), interactive asthma treatment plan, and gamification features	AP[b]/ iTouch	100% of the participants used it once daily, with 81% of the participants using it multiples times a day for a period of 30 days
Cushing et al, 2016 [32]	Asthmahero	NR	Medication usage and reminders, medication adherence graphics	AP mobile/tablet	NR
Goyal et al, 2017 [25]	"bant"	1-hour tutorial	Automatic Data Transfer, Electronic Logbook, Trends, Trend Wizard, Reward System, banter, and Personal Health Record	AP/mobile	35% of the participants had a moderate[c] or high use
Leonard et al, 2017 [33]	ITP[d] app	Day 0 of a 90-day program	"real time" adherence tracking, education modules, and patient support (reminder alert messages and behavioral reinforcement)	AP mobile/tablet	81% of the participants used it daily for a period of 90 days
Castensøe -Seidenfaden et al, 2018 [34]	Young with Diabetes	10-min tutorial	My Page, My Department, Chat Room, Carbohydrate Counting, Information About, Tips Package, To Parents, and Reminder Function	AP and AN[e] mobile/tablet	70% of the participants used it for at least 5 days out of 64 days
Le Marne et al, 2018 [16]	EpApp	Download appointment	Patients' epilepsy profiles, medication reminders, seizure diary, and personalized seizure statistics and graphs	AP and AN mobile/tablet	23% of the participants used it daily for 28 days

[a]NR: not recorded.

[b]AP: Apple.

[c]Moderate: data upload less than 3 of 7 days; High: data upload ≥3 of 7 days.

[d]ITP: Intensive Training Program app.

[e]AN: Android.

Table 3. Study outcomes.

Author, year	Setting	Mention of transition in manuscript	Significant outcomes[a]
Farooqui et al, 2015 [31]	Outpatient Clinic Ohio State University and Wexner Medical Center, United States	No	Increased treatment adherence (85%) and avoidance of asthma triggers (69%). All participants reported better knowledge of asthma after using the app.
Cushing et al, 2016 [32]	Mount Sinai Hospital Pediatric Outpatient Clinics (New York, United States)	No	Increased treatment adherence and asthma control confidence improved through the use of mobile app message reminders that led to changes in the medication use routine.
Goyal et al, 2017 [25]	The Hospital for Sick Children Toronto and University of Toronto, Canada	No	No significant differences: glycated hemoglobin, impact on self-management.
Leonard et al, 2017 [33]	Department of Pediatrics Duke University, North Carolina, United States	No	Increased treatment adherence (80%) and knowledge retention (96%). Clinically relevant decrease in serum ferritin at 6 months (P=.07).
Castensøe -Seidenfaden et al, 2018 [34]	Nordsjælland, Hervel, Roskilde, and Køge Hospitals and Steno Diabetes Center Copenhagen Denmark	Abstract, keywords, and Conclusions	No significant differences: glycated hemoglobin, impact on self-management.
Le Marne et al, 2018 [16]	The Sydney Children's Hospitals Network at Sydney, Australia	Introduction and Discussion	Increased treatment adherence (P=.045). Increased knowledge (P≤.005). No significant improvement in seizure burden or psychosocial measures.

[a]Results were statistically significant if P≤.05.

User Input Preapp Development

A total of 3 studies tested the preliminary app design and incorporated the adolescents' feedback in the most recent design of the app [16,25,34]. The design principles for the "bant app" were derived from thematic analysis of interviews conducted with adolescents who had type 1 diabetes, as well as their parents [20]. A total of 20 adolescents aged 12 to 16 years, with type 1 diabetes, were then recruited for a 12-week evaluation phase. The pilot trial study showed that daily average frequency

of blood glucose measurements increased by 50% (P=.006). HbA_{1c} did not change significantly (P=.11). Satisfaction with the "bant" app was high, with 88% of the participants stating that they would continue to use the app.

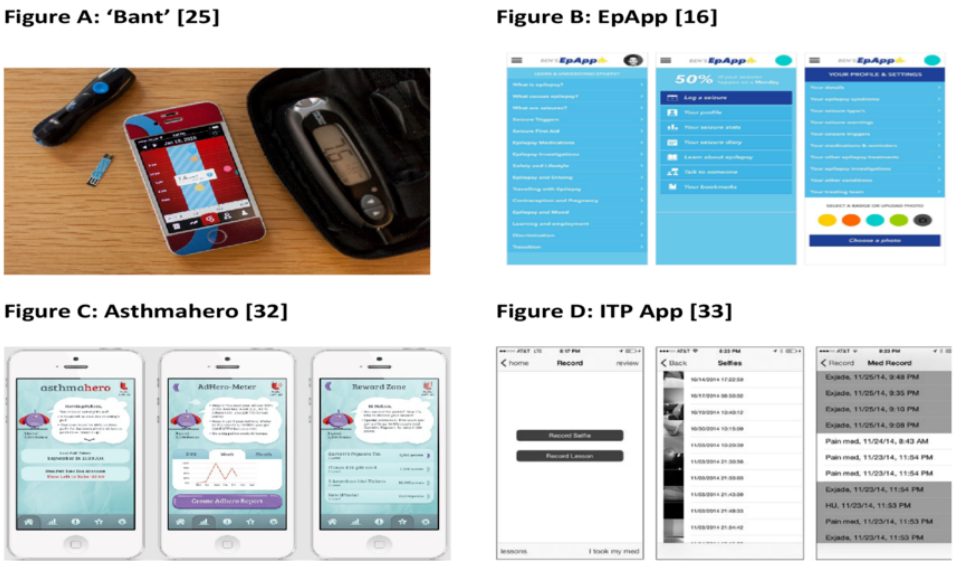
Workshops, mail panel, and feedbacks, including young people with type 1 diabetes, their parents, and health providers, were performed for the development of the "Young with Diabetes" app [35]. A feasibility study was conducted for 5 weeks among health care providers and young people. Participants found the app helpful, providing them with a range of self-management supports, such as the opportunity to write to their health care providers. They all reported that they would recommend the app to peers. Health care providers described the app as both intuitive to use and relevant to collaborating with young people with type 1 diabetes mellitus.

Le Marne et al obtained feedback for the "EpApp" on preliminary design concepts, planned features, usability, draft educational content, and potential app names through initial focus groups composed by adolescents with epilepsy and their parents, which was evaluated in a prospective cohort study [16].

The Apps

A total of 4 of the included studies provided a screenshot of the mobile app used as an intervention, and these screen shots are reproduced as Figure 2. App features are summarized in Table 2.

Figure 2. Screenshots of 4 apps.



Goyal et al [25] and Castensøe-Seidenfaden et al [34] used apps targeted at improving type 1 diabetes control. Le Marne et al's app was for education and self-management of adolescents with epilepsy [16]. Cushing et al [32] and Farooqui et al [31] targeted asthma self-management skills, and Leonard et al addressed beta thalassemia major and sickle cell disease and chelation therapy [33].

The participants of 4 out of the 6 studies [16,25,33,34] had some app training before the use of the app (Table 2). Relevant concepts and medical terminologies [16], hardware setup, introduction to app features, username creation, troubleshooting step for potential issue [25], guidance session [34], and initial education module about iron overload and chelation therapy [33] were provided as part of this training.

Table 2 provides details on the interactive features of the different apps. All of the apps allowed participants to set medication reminders. A total of 5 apps came with a feature that graphically displays medication adherence [16,25,31,33,34], whereas one of the apps measured medication adherence through "selfie" video recording [33]. A total of 4 of the apps provided disease-related information [16,34,31,33], and 1 app also provided information on specific topics such as sex, alcohol, and holding a driver's license having diabetes [34]. A total of 2 apps included a feature for social media interaction with peers [25,34]. The Intensive Training Program app used by Leonard et al allowed providers to send weekly messages to the patients, providing feedback and reinforcing the participants' adherence behaviors [33]. A total of 3 of the apps provided rewards to encourage self-management behaviors [25,31,32]. The "bant" app participants [25] could redeem their points for iTunes gift cards, whereas the AsthmaCare [31] and Asthmahero [32] apps used gamification features with rewarding points to encourage medication adherence.

A total of 5 of the 6 studies reported frequency of app usage, but the grading of frequency varied among studies, as did the measurement chosen for usage (Table 2). For instance, Farooqui et al measured frequency in terms of electronic logging of medication use [31], whereas Leonard et al measured frequency in terms of data transfer (selfie videos) [33]. The grading of frequency also varied and generally relied on data upload frequency. Goyal et al reported the app use on the basis of the total number of days that the participants wirelessly uploaded blood glucose readings to the app over the 12 months [25]. They

used 4 levels of engagement: very low, low, moderate, and high. "Young with Diabetes" use was defined as using the app for at least 5 out of 64 days [34]. On the contrary, Le Marne et al categorized app use as "daily," "weekly," "monthly," "occasionally," or "never" [16].

There were a variety of features relevant to convenience and privacy. For example, 2 studies [25,31] required the participant to carry a device other than the participant's own smartphone (iPhone or iPod Touch). This was provided as part of the intervention. A total of 3 studies [16,33,34] allowed the participants' parents to access the app, whereas 2 studies allowed health care professionals to [33,34] access the app. Only 2 studies mentioned the word transition [16,34]. Castensøe-Seidenfaden et al mentioned that "young people often struggle to self-manage type 1 diabetes during the transition from childhood to adulthood" and that "health care providers should routinely address sensitive topics and be aware of parents' need for guidance as to how to effectively support their child during the transition from childhood to adulthood" [34]. Le Marne et al stated that "self-management tools have the potential to help scaffold adolescents as they transition to adulthood" and that "developing independence and responsibility for medication is an important step in transitioning from adolescence to adulthood" [16]. None of the included studies assessed effective transition from pediatric to adult health care services as an outcome.

A total of 1 study made a reference to cost, the app being freely available via Android and Apple platforms [16]. Technical specifications varied across the apps, as did reports of problems encountered. A total of 3 studies [16,33,34] provided an app access code that permitted analytics on app usage. A total of 3 studies reported on technical issues: faulty alerts on Android devices [16], unable to open the app (Apple), unable to upload photos (Android), app not opening because of update and reinstallation needed (Apple) [34], and video recording problems [33].

Outcomes

Table 3 presents the outcome data for the 6 studies. Neither of the type 1 diabetes mellitus studies showed improvement in HbA_{1c} or self-management [25,34]. Exploratory analysis by Goyal et al showed that users who had more frequent self-monitoring of blood glucose had improvements in HbA_{1c} [25]. Castensøe-Seidenfaden et al concluded that the app might be a useful tool to complement self-management in adolescents with type 1 diabetes [34]. The other 4 studies reported increased knowledge of the condition and increased therapy adherence, including increased medication adherence through the use of "selfie videos" of the participant receiving daily chelation therapy [16,31-33].

Participants' satisfaction was reported for all studies. A total of 3 quarters of the participants were very satisfied with the "bant" app [25] and 96% of the participants reported that they would continue to use the app if it were available to them outside of the trial. A total of 45% of the subjects ranked the trending feature, which reports consecutive out-of-range blood glucose readings and prompts the user to identify the likely cause and potential solution of the trend, as the most useful component of the app [25]. A total of 78% of the participants reported that the "Young with Diabetes" app was helpful, and 85 % of the participants reported they would recommend it to others [34]. The "chat room" and "my page" were the most popular features. Topics such as sex, diabetes, driver's licenses, alcohol, and parties were the most retrieved from the information feature of the app [34]. Participants in the "EpApp" study identified themes relating to information content and reminders as the most helpful app features [16]. "Asthmahero" [32] helped as a medication reminder and improved the participants' confidence about asthma control. The gaming aspects of the app served as motivation for treatment adherence. All "Asthmacare" users reported better understanding of their asthma after the intervention, and 95% of the participants preferred using an app for asthma education over other modalities [31]. Participants of the Intensive Training Program for chelation reported that the app was easy to use and endorsed the statement that continuous use of the app would be helpful for therapy adherence [33].

Quality Assessment

The quality of the included full manuscripts varied, with individual scores shown in Multimedia Appendices 2 and 3. A total of 2 of the studies were rated as being "good," 3 studies were rated to be "fair," and 1 study was rated as "poor." The mean score of the included studies was 17.3 (range 14-22). Studies scored poorly on the following quality assessment items: clearly describing the distribution of principal confounders (4/6), participant blinding (6/6), blinding of the outcome appraiser (6/6), subject randomization (4/6), and failure to adjust for confounding factors in the analysis (4/6). None of the studies described the entire participant population and as a consequence the external validity of all the studies was rated "poor." Only 2 studies reported power to detect any clinically important effect [25,34].

Discussion

Principal Findings

The first aim of this systematic review was to evaluate the utility and effectiveness of mobile or Web-based health interventions targeting adolescents with chronic physical illnesses. Despite the proliferation of health apps over the past 5 years [18], our systematic review identified only 6 new studies. This may reflect the cost and smaller potential user base for what are quite complex apps. Few studies reached statistical significance in any of their multiple chosen outcome measures despite the relevance of these outcomes. Possible reasons for this finding include inadequate power, too short a duration of use, and failure to sufficiently engage adolescent users. A total of 2 studies specifically commented on a fixed design, which did not allow tailoring to adolescent needs [25,34]. Furthermore, when significant outcomes were reported, these related to adherence, except in diabetes, the most frequent specialist-to-specialist transition [36]. The latter may be because of the intrinsic utility of the app or because of inadequate duration of the study to influence the outcome measure of HbA_{1c}.

The use of apps is promoted as having the potential to improve patterns of self-management support [37] that are necessary during the transition process to adult care. However, most studies identified in the initial searches focused on mobile or Web-based health interventions for adult self-management, rather than for adolescents. The data from adult studies likely have limited generalizability to adolescents. Adolescents have grown up in a technological era, fully embracing technology as a way to interact with others and becoming both expert and selective in electronic media and on social networking sites [38]. Apps developed for adults tend to be more directive [39,40], as opposed to delivering features that might engage young people [41]. Only 3 studies reported adolescents in predevelopment design, but it was difficult to assess if this codesign had an impact on outcomes.

The secondary aim of the study was to describe features of the apps that might explicitly or implicitly facilitate the transition of adolescents with chronic physical illness from pediatric to adult care. We were unable to draw any conclusions related to this aim. None of the studies measured transition process as an outcome; indeed, there are no established, validated successful outcome criteria for the transition process of adolescent and young adults with chronic illness [42]. In addition, the duration of most of the included studies was less than 6 months, which would not provide an adequate length of time to measure transition outcomes.

Education was a feature in 4 apps [16,31,33,34]. Educating the patient is an important aspect for the development of self-management skills. The ongoing accessibility of illness-specific educational themes in an electronic device allows patients to easily obtain reliable information. Nevertheless, education alone is insufficient to increase medication adherence and the development of self-management skills [9,43], and adding features, such as reminders, goal setting, rewards, and social media interaction with peers, may enhance better outcomes [43].

A total of 2 apps [25,34] included a feature that allowed the participants to share their experiences with other peers through social media to complement education and reassurance of common lived experiences. In addition, 2 apps [31,32] included gamification features, and 5 apps included graphical displays of illness-specific control measures. Patients liked the gaming aspects of these mobile apps and felt that earning points served as motivation for medication adherence and continuous app use. This feature would add considerably to future app development. Knowing what features patients like in the apps might assist in the development of a generic rather than an illness-specific app that could incorporate these identified common features and be easily tailored for purpose through simple “in-app” modifications.

No study presented an economic evaluation. Even though mobile health interventions are often described as cost-effective tools for patients living with chronic illnesses, there are limited economic data to support this [44]. A total of 2 systematic reviews have been conducted to assess the literature on the economic value of mobile health interventions. Both reviews concluded that there was a lack of economic data to support mobile health interventions for patients with chronic illness [45,46]. Iribarren et al suggested that studies should follow the established economic reporting guidelines to improve the available data [45], whereas Badawy and Kuhns highlighted the need for comprehensive economic evaluation of these interventions to understand the association between their cost-effectiveness while supporting self-management of patients with chronic illnesses [46].

A limitation of our review was the small number and heterogeneity of the studies that precluded meta-analysis. The lack of data about app development strategies that included adolescent input is an important limitation around feasibility and utility. The strength of this review is that it provides an up-to-date overview of apps as these relate to health care in adolescents with chronic illness. The small number of studies identified allowed us to describe apps in more detail, which may be helpful to those considering app development.

Conclusions

In conclusion, this review has found that there remain limited data about the utility and effectiveness of mobile and Web-based apps facilitating self-management and the transition of adolescents with chronic illness from pediatric to adult health care services. The latter is a key time when independent adolescent-friendly assistance is required. This potentially reflects the lack of commercial appeal of such apps and the high cost of developing apps that compete with the thousands available to young people. Future studies could consider cocreation with adolescents, financial assistance from lay illness support groups, and easy but confidential ways to report feedback through the app.

Acknowledgments

The authors would like to thank Ms Trish Bennett, Library Manager at The Children’s Hospital at Westmead Medical Library, for her assistance in database searching optimization.

Authors' Contributions

YIVP developed and conducted search strategy, designed extraction template, extracted data, drafted the manuscript, and revised subsequent drafts. SM and JH critically reviewed the manuscript. KS developed search strategy, designed extraction template, extracted data, contributed to the writing of the manuscript, and critically reviewed and revised draft manuscripts. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy Medical Literature Analysis and Retrieval System Online via OvidSP (1946-present).
[PDF File (Adobe PDF File), 55 KB - jmir_v21i11e13579_app1.pdf]

Multimedia Appendix 2
Description of items and scores used for the quality assessment of the included studies.
[PDF File (Adobe PDF File), 58 KB - jmir_v21i11e13579_app2.pdf]

Multimedia Appendix 3
Designated values for the quality assessment of the included studies.
[PDF File (Adobe PDF File), 54 KB - jmir_v21i11e13579_app3.pdf]

References

1. Rosina R, Crisp J, Steinbeck K. Treatment adherence of youth and young adults with and without a chronic illness. Nurs Health Sci 2003 Jun;5(2):139-147. [doi: 10.1046/j.1442-2018.2003.00149.x] [Medline: 12709169]
2. van Cleave J, Gortmaker SL, Perrin JM. Dynamics of obesity and chronic health conditions among children and youth. J Am Med Assoc 2010 Feb 17;303(7):623-630. [doi: 10.1001/jama.2010.104] [Medline: 20159870]
3. Perrin JM, Bloom SR, Gortmaker SL. The increase of childhood chronic conditions in the United States. J Am Med Assoc 2007 Jun 27;297(24):2755-2759. [doi: 10.1001/jama.297.24.2755] [Medline: 17595277]
4. Steinbeck K, Brodie L, Towns S. Chronic illness and transition to adult care. In: Bennett DL, Towns SJ, Elliott EJ, Merrick J, editors. Challenges in Adolescent Health. An Australian Perspective. First Edition. New York: Nova Science Publishers Inc; 2009:225-234.
5. Australian Institute of Health and Welfare. 2018. Australia's Health 2018 URL: https://www.aihw.gov.au/getmedia/7c42913d-295f-4bc9-9c24-4e44eff4a04a/aihw-aus-221.pdf.aspx?inline=true [accessed 2018-10-30]
6. Barlow J, Wright C, Sheasby J, Turner A, Hainsworth J. Self-management approaches for people with chronic conditions: a review. Patient Educ Couns 2002;48(2):177-187. [doi: 10.1016/s0738-3991(02)00032-0] [Medline: 12401421]
7. Lorig KR, Holman H. Self-management education: history, definition, outcomes, and mechanisms. Ann Behav Med 2003 Aug;26(1):1-7. [doi: 10.1207/S15324796ABM2601_01] [Medline: 12867348]
8. Huang JS, Terrones L, Tompane T, Dillon L, Pian M, Gottschalk M, et al. Preparing adolescents with chronic disease for transition to adult care: a technology program. Pediatrics 2014 Jun;133(6):e1639-e1646 [FREE Full text] [doi: 10.1542/peds.2013-2830] [Medline: 24843066]
9. Pai A, Ostendorf HM. Treatment adherence in adolescents and young adults affected by chronic illness during the health care transition from pediatric to adult health care: a literature review. Child Heal Care 2011 Jan;40(1):16-33. [doi: 10.1080/02739615.2011.537934]
10. Crowley R, Wolfe I, Lock K, McKee M. Improving the transition between paediatric and adult healthcare: a systematic review. Arch Dis Child 2011 Jun;96(6):548-553. [doi: 10.1136/adc.2010.202473] [Medline: 21388969]
11. Campbell F, Biggs K, Aldiss SK, O'Neill PM, Clowes M, McDonagh J, et al. Transition of care for adolescents from paediatric services to adult health services. Cochrane Database Syst Rev 2016 Apr 29;4:CD009794. [doi: 10.1002/14651858.CD009794.pub2] [Medline: 27128768]
12. Viner RM. Transition of care from paediatric to adult services: one part of improved health services for adolescents. Arch Dis Child 2008 Feb;93(2):160-163. [doi: 10.1136/adc.2006.103721] [Medline: 17942588]
13. Pew Research Center. 2018. Teens, Social Media & Technology 2018 URL: https://www.pewinternet.org/2018/05/31/teens-social-media-technology-2018/ [accessed 2019-05-25]
14. Ofcom. 2014 Aug 7. The Communications Market Report URL: https://www.ofcom.org.uk/__data/assets/pdf_file/0031/19498/2014_uk_cmr.pdf [accessed 2019-02-15]
15. Australian Communications and Media Authority. 2013. Mobile Apps: Putting the ‘smart’ in Smartphones URL: https://www.acma.gov.au/theACMA/engage-blogs/engage-blogs/Research-snapshots/Mobile-apps-putting-the-smart-in-smartphones [accessed 2019-01-31]
16. Le Marne FA, Butler S, Beavis E, Gill D, Bye AM. EpApp: Development and evaluation of a smartphone/tablet app for adolescents with epilepsy. J Clin Neurosci 2018 Apr;50:214-220. [doi: 10.1016/j.jocn.2018.01.065] [Medline: 29422360]
17. Dennison L, Morrison L, Conway G, Yardley L. Opportunities and challenges for smartphone applications in supporting health behavior change: qualitative study. J Med Internet Res 2013 Apr;15(4):e86 [FREE Full text] [doi: 10.2196/jmir.2583] [Medline: 23598614]
18. Peng W, Kanthawala S, Yuan S, Hussain SA. A qualitative study of user perceptions of mobile health apps. BMC Public Health 2016 Nov 14;16(1):1158 [FREE Full text] [doi: 10.1186/s12889-016-3808-0] [Medline: 27842533]

19. Majeed-Ariss R, Baildam E, Campbell M, Chieng A, Fallon D, Hall A, et al. Apps and adolescents: a systematic review of adolescents' use of mobile phone and tablet apps that support personal management of their chronic or long-term physical conditions. J Med Internet Res 2015 Dec 23;17(12):e287 [FREE Full text] [doi: 10.2196/jmir.5043] [Medline: 26701961]
20. Cafazzo JA, Casselman M, Hamming N, Katzman DK, Palmert MR. Design of an mHealth app for the self-management of adolescent type 1 diabetes: a pilot study. J Med Internet Res 2012 May 8;14(3):e70 [FREE Full text] [doi: 10.2196/jmir.2058] [Medline: 22564332]
21. Frøisland DH, Arsand E, Skårderud F. Improving diabetes care for young people with type 1 diabetes through visual learning on mobile phones: mixed-methods study. J Med Internet Res 2012 Aug 6;14(4):e111 [FREE Full text] [doi: 10.2196/jmir.2155] [Medline: 22868871]
22. Burbank AJ, Lewis SD, Hewes M, Schellhase DE, Rettiganti M, Hall-Barrow J, et al. Mobile-based asthma action plans for adolescents. J Asthma 2015;52(6):583-586 [FREE Full text] [doi: 10.3109/02770903.2014.995307] [Medline: 25494553]
23. Aldiss S, Taylor RM, Soanes L, Maguire R, Sage M, Kearney N, et al. Working in collaboration with young people and health professionals. A staged approach to the implementation of a randomised controlled trial. J Res Nurs 2010 Sep 17;16(6):561-576 [FREE Full text] [doi: 10.1177/1744987110380803]
24. Fedele DA, Cushing CC, Fritz A, Amaro CM, Ortega A. Mobile health interventions for improving health outcomes in youth: a meta-analysis. JAMA Pediatr 2017 May;171(5):461-469 [FREE Full text] [doi: 10.1001/jamapediatrics.2017.0042] [Medline: 28319239]
25. Goyal S, Nunn CA, Rotondi M, Couperthwaite AB, Reiser S, Simone A, et al. A mobile app for the self-management of type 1 diabetes among adolescents: a randomized controlled trial. JMIR Mhealth Uhealth 2017 Jun 19;5(6):e82 [FREE Full text] [doi: 10.2196/mhealth.7336] [Medline: 28630037]
26. World Health Organization. Maternal, newborn, child and adolescent health URL: http://www.who.int/topics/adolescent_health/en/ [accessed 2018-11-28]
27. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Br Med J 2009 Jul 21;339:b2535 [FREE Full text] [doi: 10.1136/bmj.b2535] [Medline: 19622551]
28. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. J Epidemiol Community Health 1998 Jun;52(6):377-384 [FREE Full text] [doi: 10.1136/jech.52.6.377] [Medline: 9764259]
29. Sterling MR, Shaw AL, Leung PB, Safford MM, Jones CD, Tsui EK, et al. Home care workers in heart failure: a systematic review. J Multidiscip Healthc 2018;11:481-492 [FREE Full text] [doi: 10.2147/JMDH.S175512] [Medline: 30288046]
30. Samoocha D, Bruinvels DJ, Elbers NA, Anema JR, van der Beek AJ. Effectiveness of web-based interventions on patient empowerment: a systematic review and meta-analysis. J Med Internet Res 2010 Jun 24;12(2):e23 [FREE Full text] [doi: 10.2196/jmir.1286] [Medline: 20581001]
31. Farooqui N, Phillips G, Barrett C, Stukus D. Acceptability of an interactive asthma management mobile health application for children and adolescents. Ann Allergy Asthma Immunol 2015 Jun;114(6):527-529. [doi: 10.1016/j.anai.2015.03.006] [Medline: 26021895]
32. Cushing A, Manice MP, Ting A, Parides MK. Feasibility of a novel mHealth management system to capture and improve medication adherence among adolescents with asthma. Patient Prefer Adherence 2016;10:2271-2275 [FREE Full text] [doi: 10.2147/PPA.S115713] [Medline: 27853357]
33. Leonard S, Anderson LM, Jonassaint J, Jonassaint C, Shah N. Utilizing a novel mobile health 'selfie' application to improve compliance to iron chelation in pediatric patients receiving chronic transfusions. J Pediatr Hematol Oncol 2017 Apr;39(3):223-229. [doi: 10.1097/MPH.0000000000000743] [Medline: 28099398]
34. Castensøe-Seidenfaden P, Husted GR, Jensen AK, Hommel E, Olsen B, Pedersen-Bjergaard U, et al. Testing a smartphone app (young with diabetes) to improve self-management of diabetes over 12 months: randomized controlled trial. JMIR Mhealth Uhealth 2018 Jun 26;6(6):e141 [FREE Full text] [doi: 10.2196/mhealth.9487] [Medline: 29945861]
35. Castensøe-Seidenfaden P, Husted G, Teilmann G, Hommel E, Olsen BS, Kensing F. Designing a self-management app for young people with type 1 diabetes: methodological challenges, experiences, and recommendations. JMIR Mhealth Uhealth 2017 Oct 23;5(10):e124 [FREE Full text] [doi: 10.2196/mhealth.8137] [Medline: 29061552]
36. Steinbeck KS, Brodie L, Towns SJ. Transition in chronic illness: who is going where? J Paediatr Child Health 2008 Sep;44(9):478-482. [doi: 10.1111/j.1440-1754.2008.01321.x] [Medline: 18928466]
37. Husted GR, Weis J, Teilmann G, Castensøe-Seidenfaden P. Exploring the influence of a smartphone app (young with diabetes) on young people's self-management: qualitative study. JMIR Mhealth Uhealth 2018 Feb 28;6(2):e43 [FREE Full text] [doi: 10.2196/mhealth.8876] [Medline: 29490897]
38. Spies Shapiro LA, Margolin G. Growing up wired: social networking sites and adolescent psychosocial development. Clin Child Fam Psychol Rev 2014 Mar;17(1):1-18 [FREE Full text] [doi: 10.1007/s10567-013-0135-1] [Medline: 23645343]
39. Fu H, McMahon SK, Gross CR, Adam TJ, Wyman JF. Usability and clinical efficacy of diabetes mobile applications for adults with type 2 diabetes: A systematic review. Diabetes Res Clin Pract 2017 Sep;131:70-81. [doi: 10.1016/j.diabres.2017.06.016] [Medline: 28692830]

40. Veazie S, Winchell K, Gilbert J, Paynter R, Ivlev I, Eden KB, et al. Rapid evidence review of mobile applications for self-management of diabetes. J Gen Intern Med 2018 Jul;33(7):1167-1176 [FREE Full text] [doi: 10.1007/s11606-018-4410-1] [Medline: 29740786]
41. Badawy SM, Thompson AA, Kuhns LM. Medication adherence and technology-based interventions for adolescents with chronic health conditions: a few key considerations. JMIR Mhealth Uhealth 2017 Dec 22;5(12):e202 [FREE Full text] [doi: 10.2196/mhealth.8310] [Medline: 29273573]
42. Fair C, Cuttance J, Sharma N, Maslow G, Wiener L, Betz C, International and Interdisciplinary Health Care Transition Research Consortium. International and interdisciplinary identification of health care transition outcomes. JAMA Pediatr 2016 Mar;170(3):205-211 [FREE Full text] [doi: 10.1001/jamapediatrics.2015.3168] [Medline: 26619178]
43. Dean AJ, Walters J, Hall A. A systematic review of interventions to enhance medication adherence in children and adolescents with chronic illness. Arch Dis Child 2010 Sep;95(9):717-723. [doi: 10.1136/adc.2009.175125] [Medline: 20522463]
44. Schweitzer J, Synowiec C. The economics of eHealth and mHealth. J Health Commun 2012;17(Suppl 1):73-81. [doi: 10.1080/10810730.2011.649158] [Medline: 22548602]
45. Iribarren SJ, Cato K, Falzon L, Stone PW. What is the economic evidence for mHealth? A systematic review of economic evaluations of mHealth solutions. PLoS One 2017;12(2):e0170581 [FREE Full text] [doi: 10.1371/journal.pone.0170581] [Medline: 28152012]
46. Badawy SM, Kuhns LM. Economic evaluation of text-messaging and smartphone-based interventions to improve medication adherence in adolescents with chronic health conditions: a systematic review. JMIR Mhealth Uhealth 2016 Oct 25;4(4):e121 [FREE Full text] [doi: 10.2196/mhealth.6425] [Medline: 27780795]

Abbreviations

HbA_{1c}: glycated hemoglobin
MEDLINE: Medical Literature Analysis and Retrieval System Online
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by G Eysenbach; submitted 01.02.19; peer-reviewed by T Lakiang, C Jonassaint, S Badawy; comments to author 28.04.19; revised version received 19.06.19; accepted 07.07.19; published 20.11.19.

Please cite as:
Virella Pérez YI, Medlow S, Ho J, Steinbeck K
Mobile and Web-Based Apps That Support Self-Management and Transition in Young People With Chronic Illness: Systematic Review
J Med Internet Res 2019;21(11):e13579
URL: http://www.jmir.org/2019/11/e13579/
doi:10.2196/13579
PMID:31746773

Original Paper

Mechanisms and Effects of a WeChat-Based Intervention on Suicide Among People Living With HIV and Depression: Path Model Analysis of a Randomized Controlled Trial

Yiran Li[1*], MB; Yan Guo[1,2,3*], PhD; Y Alicia Hong[4*], PhD; Mengting Zhu[1*], BS; Chengbo Zeng[5,6], MS; Jiaying Qiao[1], MS; Zhimeng Xu[1], MS; Hanxi Zhang[7], MS; Yu Zeng[1], MB; Weiping Cai[8], MD; Linghua Li[8], MD; Cong Liu[8], MSN

[1]Department of Epidemiology and Biostatistics, School of Public Health, Sun Yat-sen University, Guangzhou, China

[2]Sun Yat-sen Center for Migrant Health Policy, Guangzhou, China

[3]Sun Yat-sen Center for Global Health, Institute of State Governance, Guangzhou, China

[4]Department of Health Administration and Policy, College of Health and Human Services, George Mason University, Fairfax, VA, United States

[5]South Carolina SmartState Center for Healthcare Quality, Arnold School of Public Health, University of South Carolina, Columbia, SC, United States

[6]Department of Health Promotion, Education, and Behavior, Arnold School of Public Health, University of South Carolina, Columbia, SC, United States

[7]National Center of AIDS/STD Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing, China

[8]Department of Infectious Diseases, Eight People's Hospital, Guangzhou, China

[*]these authors contributed equally

Corresponding Author:
Yan Guo, PhD
Department of Epidemiology and Biostatistics
School of Public Health
Sun Yat-sen University
#74 Zhongshan 2nd Road
Guangzhou,
China
Phone: 86 020 87334202
Email: guoy8@mail.sysu.edu.cn

Abstract

Background: People living with HIV and depression have high rates of suicide. Studies of mobile health (mHealth) interventions have shown feasibility, acceptability, and efficacy in improving mental health in people living with HIV and depression. However, few studies have examined the mechanisms and effects of mHealth interventions on suicide.

Objective: This study was designed to examine the mechanisms and effects of a WeChat-based intervention, *Run4Love*, on suicide among people living with HIV and depression in China, while considering perceived stress and depressive symptoms as mediators.

Methods: A sample of 300 People living with HIV and depression was recruited from the outpatient clinic of a large HIV or AIDS treatment hospital and was randomized to the *Run4Love* group or a control group. Data were collected at baseline, 3-, 6-, and 9-month follow-ups. Path analysis modeling, with longitudinal data, was used in data analyses.

Results: The *Run4Love* mHealth intervention had a direct effect on reducing suicide rate at the 6-month follow-up (beta=−.18, *P*=.02) and indirect effect through reducing perceived stress and depressive symptoms at the 3-month follow-up (beta=−.09, *P*=.001). A partial mediating effect between perceived stress and depressive symptoms accounted for 33% (–0.09/–0.27) of the total effect.

Conclusions: Through path analyses, we understood the mechanisms and effects of an mHealth intervention on suicide prevention. The findings underscored the importance of stress reduction and depression treatment in such a program. We call for more effective suicide prevention, especially mHealth interventions targeting the vulnerable population of people living with HIV and depression.

Trial Registration: Chinese Clinical Trial Registry ChiCTR-IPR-17012606; http://www.chictr.org.cn/showprojen.aspx?proj=21019

(J Med Internet Res 2019;21(11):e14729) doi:10.2196/14729

KEYWORDS

HIV; mHealth; depression; suicide

Introduction

Background

People living with HIV have high rates of depressive symptoms. For example, a nationally representative survey in the United States reported that 36% of people living with HIV had depression [1]. In China, the rate of depression among people living with HIV was 50.8%, according to a recent review study [2]. Out of 37.9 million people living with HIV in the world [3], more than 12 million are living with depression [4]. The literature has shown a causal relationship between depression and suicide [5,6]. Previous studies have also reported high suicide rates in people living with HIV. A survey of 1560 people living with HIV in the United States revealed that 26% of the participants had a suicidal ideation and 13% had a suicidal attempt [7]. According to a recent survey in China, 32.4% of people living with HIV had had suicidal ideation or behavior since HIV diagnosis [8]. Such high rates of suicide and depression warrant effective interventions targeting the vulnerable population of people living with HIV and depression, especially the interventions that can reach a large number of people living with HIV and depression efficiently, such as mobile health (mHealth) interventions.

Reviews

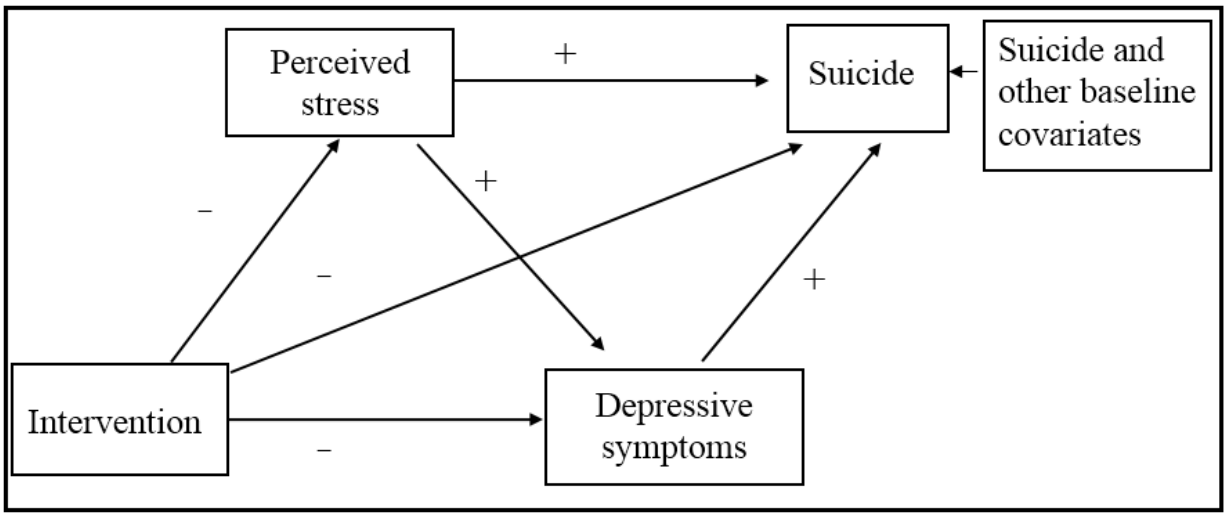
According to a recent systematic review, the majority of the existing mHealth interventions targeting people living with HIV were focused on medication adherence or retention in care [9]. Only a small number of mHealth interventions were implemented to improve mental health outcomes or decrease suicidal risks among people living with HIV and depression [9-11]. In a pilot mHealth intervention, Swendeman and colleagues developed a self-monitoring program, which showed initial efficacy in reducing stress in people living with HIV [12]. Van Luenen and colleagues tested a Web-based cognitive behavioral therapy program via a randomized controlled trial (RCT) with 188 people living with HIV and depression. The intervention successfully reduced depressive symptoms and anxiety of the intervention group at 8-week and 12-week follow-ups compared with the control group [13]. Despite the growing interest and initial efficacy of mHealth interventions to deliver mental health services and prevent suicide among people living with HIV and depression, no study existed on the potential mechanism of the intervention in mental health outcomes, especially on the basis of a longitudinal design.

Aims

Accordingly, in this study, we aimed to examine the mechanisms of the mHealth intervention *Run4Love* in a suicide outcome in a sample of people living with HIV and depression by conducting path model analyses, using longitudinal data from an RCT [14]. The *Run4Love* was a WeChat-based intervention, comprising 2 major components: the adapted cognitive-behavioral stress management (CBSM) course and physical activity promotion. It was designed to reduce stress through multiple coping strategies.

The literature has documented the importance of stress reduction in depression prevention [15-18]. Previous studies also showed that depressive symptoms mediated perceived stress and suicidal behaviors [19-21]. However, no study has investigated the mechanisms of stress and depressive symptoms in suicide in the context of mHealth intervention. On the basis of the associations among perceived stress, depressive symptoms, and suicide illustrated in the previous studies [19-21], we hypothesized that perceived stress and depressive symptoms at a 3-month follow-up served as mediators on the effects of the mHealth intervention on suicide at a 6-month follow-up. The hypothesized model is depicted in Figure 1.

Figure 1. Hypothesized path model of intervention, perceived stress, depressive symptoms, and suicide in people living with HIV and depression.



Methods

Research Setting

A total of 300 participants were recruited from an outpatient clinic of the only designated hospital for HIV or AIDS treatment in Guangzhou, China. As the capital city of Guangdong Province, Guangzhou is the third largest city in China. According to the AIDS Prevention and Treatment Report in 2017, more than 53,600 people living with HIV live in Guangdong province [22].

Participant Eligibility and Recruitment

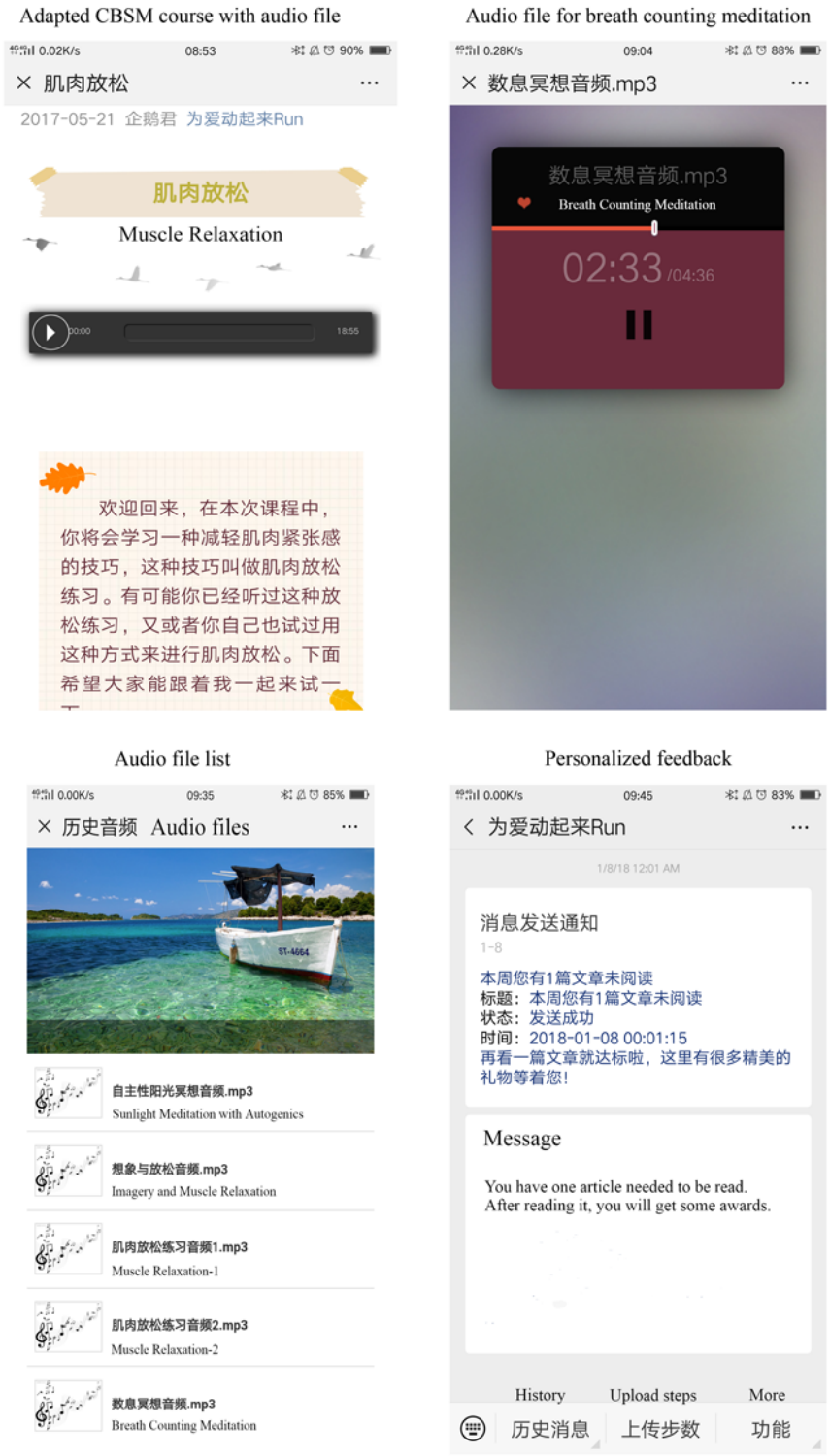
We recruited participants in the waiting room of an outpatient clinic. A member of the research team invited patients to participate in a research study. Those who showed initial interest were invited to a private space for further explanation and screening. The participation criteria included the following: (1) at least 18 years old, (2) HIV-seropositive status (registered in the hospital system or with an official document), (3) having clinically significant depressive symptoms (Center for Epidemiologic Studies-Depression; CES-D score≥16), and (4) having a mobile phone and WeChat account. Patients were excluded if they were (1) currently under psychiatric or psychological treatment, (2) unable to finish the questionnaire, (3) unable to read or listen to the materials on WeChat, and (4) unable to engage in physical activities because of medical reasons.

Participants who met the eligibility criteria and provided written informed consent were invited to participate. They completed a baseline survey before they were randomized to the intervention group or control group of the *Run4Love* trial. All participants received breakfast (milk and bread) as an incentive. The study protocol was approved by the Institutional Review Board of the Sun Yat-sen University.

The *Run4Love* Intervention

The *Run4Love* intervention was adapted from the CBSM program [23], which was designed as a face-to-face intervention program for people living with HIV, and it has demonstrated good efficacy in various populations and settings [24-26]. We adapted the CBSM into a multimedia format and delivered it via WeChat, the most popular social media platform in China, with over 1 billion active users [27]. The intervention protocol is detailed elsewhere [14]. Briefly, participants assigned to the intervention group received a series of adapted CBSM courses, and they were encouraged to do regular physical activities for 3 months. Multimedia information, with automatic progress tracking and personalized feedback, was delivered via the *Run4Love* WeChat account (Figure 2). The main purpose of the program was to alleviate people living with HIV's depressive symptoms and improve their quality of life by reducing perceived stress. Participants in the wait-list control group received a brochure on HIV-related nutrition. Outcomes were assessed at baseline, 3, 6, and 9 months. All participants in the control group would receive the intervention after the completion of the trial.

Figure 2. WeChat users' interfaces in the Run4Love Program. CBSM: cognitive-behavioral stress management.



Measurements

Demographic Characteristics

Participants' demographic characteristics assessed in the study included age, gender, sexual orientation, marital status, education, and duration since HIV diagnosis.

Depressive Symptoms

Depressive symptoms were assessed using the Chinese version of the CES-D scale at baseline and each follow-up [28]. The scale has been validated in various Chinese chronic patients, and it demonstrated high levels of validity and reliability [29,30]. The CES-D comprises 20 items, such as *My appetite was poor* and *I could not shake off the blues.* Participants were asked to select the frequencies of physiological symptoms during the past week. The total scores ranged from 0 to 60, with a higher

score indicating a higher level of depressive symptoms. Participants with scores of 16 or higher were considered as having clinically significant depressive symptoms [31]. The Cronbach alpha of the scale was .77 in this study.

Perceived Stress

Perceived stress was assessed with the 10-item Chinese version of Perceived Stress Scale (PSS-10) at baseline and each follow-up [32]. The scale was the most widely used instrument to measure perception of stress; the validity and reliability of the scale had been validated and established in the Chinese population [33-35]. Participants were asked about their feelings and thoughts in the past month (eg, *How often have you felt nervous and stressed*?). The total scores of the scale ranged from 0 to 40, with higher scores indicating higher levels of perceived stress. Scores ranging from 0 to 13, 14 to 26, and 27 to 40 were considered as low, moderate, and high levels of perceived stress, respectively [36]. In this study, the Cronbach alpha of the PSS-10 was .67.

Suicide

Suicide was defined as having had suicidal ideation or attempt in this study. Suicidal ideation referred to the planning or thinking of committing suicide, and suicidal attempt was the actual behavior of committing suicide. Participants were asked how many times they had had thought about committing suicide and how many times they had actually attempted suicide in the past 3 months at baseline and each follow-up. Participants who answered *yes* to either of the 2 questions were considered as having had suicide. Those who answered *no* to both of the questions were considered as not having had suicide (dummy coded as 0=not having had suicide, 1=having had suicide).

Data Analysis

First, descriptive analyses were used on demographic characteristics, perceived stress, depressive symptoms, and suicide. The continuous variables with skewed distribution (eg, age, duration since HIV infection, perceived stress, and depressive symptoms) were described using median (interquartile range; IQR), and categorical variables were described using frequencies and percentages.

Second, treatment group comparisons for the outcome and mediating variables at 3 assessment points were conducted using Wilcoxon rank-sum test for continuous variables with skewed distribution (eg, perceived stress and depressive symptoms) and Chi-square test for categorical variables (eg, suicide). Third, bivariate analyses of suicide were performed using Wilcoxon rank-sum test (for continuous variables with skewed distribution) and Chi-square test (for categorical variables).

To further examine the mechanisms of how participants' suicide changed as a result of the intervention, we conducted a path analysis after controlling for the baseline suicide and potential confounders, such as demographics. A mediation model of the path analysis was used to test the hypothesis whether the intervention effect on suicide could be explained by the mediating factors of perceived stress and depressive symptoms. Such modeling was designed to explore the mechanisms and direct and indirect effects of the intervention on suicide. The pathways of intervention→suicide, intervention→depressive symptoms→suicide, intervention→perceived stress→suicide, and intervention→perceived stress→depressive symptoms→suicide were examined, respectively (dummy coded as 0=control group, 1=intervention group). The statistical significance was defined as $P<.05$.

The model was estimated by weighted least squares means and variance–adjusted estimation. Multiple indicators were used to evaluate the goodness of fit of the model, including Chi-square statistic, Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), root mean square error of approximation (RMSEA), and weighted root mean square residual (WRMR). Smaller Chi-square value indicates better model fit. CFI≥0.95, TLI>0.90, RMSEA≤0.06, and WRMR≤1.00 indicate good model fit [37,38]. Descriptive statistics, bivariate statistics, and correlation analyses were performed using SAS version 9.4 (SAS Institute, Inc). Path analysis was performed using Mplus version 7.0 (Muthen and Muthen) [39].

Results

Baseline Characteristics of the Participants

As shown in Table 1, the median (IQR) age of the 300 participants was 27.5 (24.5, 31.3) years, with a range from 18 to 51. Most (277/300, 92.3%) of the participants were male. More than four-fifths (245/300, 81.7%) of the participants were homosexual, bisexual, or uncertain of their sexual orientations, and 87.3% (262/300) were unmarried. About 60.7% (182/300) of the participants had completed at least some college education. The median duration of HIV infection was 1.7 years (IQR 0.6, 3.8).

At baseline, a majority (256/300, 85.4%) of the participants had a moderate level of perceived stress, followed by 8.3% (25/300) having a high level and 6.3% (19/300) a low level of stress. The median (IQR) of the CES-D score was 23.0 (19.0, 28.0). Approximately 44.0% (132/300) of the participants reported having considered committing suicide and 9.7% (29/300) having tried to attempt suicide in the past 3 months. In total, about 45.0% (135/300) of the participants had suicidal ideation or attempts in the past 3 months.

The proportion of homosexual, bisexual, or sexual orientation–uncertain participants in the intervention group was slightly higher than the control group (130/150, 86.7% vs 115/150, 76.7%; $P=.03$). Other demographic characteristics, mental health outcomes (depressive symptoms and perceived stress), and suicide were balanced in the 2 groups at baseline.

Table 1. Sample characteristics of people living with HIV and depression (N=300).

Characteristics	Value
Age (years), median (IQR[a])	27.5 (24.5, 31.3)
Gender, n (%)	
Male	277 (92.3)
Female	23 (7.7)
Sexual orientation, n (%)	
Heterosexual	55 (18.3)
Homosexual/bisexual/uncertain	245 (81.7)
Education, n (%)	
≤High school	118 (39.3)
>High school	182 (60.7)
Marital status, n (%)	
Single	262 (87.3)
Married	38 (12.7)
Employment status, n (%)	
Unemployed	49 (16.3)
Employed	251 (83.7)
Duration since HIV diagnosis (years), median (IQR)	1.7 (0.6, 3.8)
Perceived stress, median (IQR)	20.0 (18.0, 23.0)
Levels of perceived stress, n (%)	
Low perceived stress	19 (6.3)
Moderate perceived stress	256 (85.3)
High perceived stress	25 (8.3)
Depressive symptoms, median (IQR)	23.0 (19.0, 28.0)
Suicidal ideation (times in the last 3 months), n (%)	
0	168 (56.0)
1-2	84 (28.0)
≥3	48 (16.0)
Suicidal attempt (times in the last 3 months), n (%)	
0	271 (90.3)
1-2	19 (6.3)
≥3	10 (3.3)
Suicide (in the last 3 months), n (%)	
Yes	135 (45.0)
No	165 (55.0)

[a]IQR: interquartile range.

Changes in Outcomes and Mediating Variables Over Time

The intervention had significant effects on the mental health outcomes over time. A total of 274 participants (91.3%, 274/300) remained at the 3-month follow-up (139 in the intervention group; 135 in the control group), and 265 (88.3%, 265/300) participants remained at the 6-month follow-up (132 in the intervention group; 133 in the control group). As reported in Table 2, the median score of the perceived stress was 20 at baseline in the intervention and control groups; it changed to 17 in the intervention group compared with 19 in the control group at the 3-month follow-up. Similarly, the depressive symptoms in the intervention group reduced from 23 at baseline to 17 at the 3-month follow-up, whereas in the control group, it remained unchanged at 23. The rate of clinically significant

depressive symptoms reduced from 100% to 56.8% (79/139) in the intervention group compared with 79.3% (107/135) in the control group at the 3-month follow-up. The rate of suicide reduced from 44.0% (66/150) at baseline to 31.7% (44/139) at the 3-month follow-up, and it further reduced to 21.2% (28/132) at the 6-month follow-up in the intervention group. The difference in the rate of suicide between the 2 groups was close to 20% (21.2% in intervention group vs 40.6% in control group) at 6 months. The intervention group had improved significantly in 3 variables—perceived stress at the 3- and 6-month follow-ups (both *P*<.001), depressive symptoms at the 3- and 6-month follow-ups (both *P*<.001), and suicide at the 6-month follow-up (*P*<.001).

Table 2. Mental health and behavior outcomes at baseline and follow-ups.

Variables	Baseline (T_0)			3 months (T_1)			6 months (T_2)		
	I[a] (n=150)	C[b] (n=150)	*P* value	I (n=139)	C (n=135)	*P* value	I (n=132)	C (n=133)	*P* value
Perceived stress, median (IQR[c])	20 (17, 22)	20 (18, 23)	.14[d]	17 (12, 20)	19 (16, 22)	<.001[d]	17 (13, 20)	19 (16, 23)	<.001[d]
Depressive symptoms, median (IQR)	23 (19, 28)	23 (19, 27)	.81[d]	17 (11, 24)	23 (18, 31)	<.001[d]	17 (10, 23)	24 (16, 32)	<.001[d]
Clinically significant depressive symptoms, n (%)	150 (100.0)	150 (100.0)	—[e]	79 (56.8)	107 (79.3)	<.001[f]	68 (51.5)	102 (76.7)	<.001[f]
Suicide, n (%)	66 (44.0)	69 (46.0)	.82[f]	44 (31.7)	49 (36.3)	.42[f]	28 (21.2)	54 (40.6)	<.001[f]

[a]Intervention group.

[b]Control group.

[c]IQR: interquartile range.

[d]Wilcoxon rank-sum test.

[e]Chi-square test is not applicable in this cell. All participants had depressive symptoms in both the intervention and control groups.

[f]Chi-square test.

Bivariate Analysis of Suicide

Results of bivariate analyses between demographic characteristics and suicide indicated that only employment was significantly associated with suicide. Specifically, employed participants reported lower rates of suicide than the unemployed (104/251, 41.4% vs 31/49, 63.3%, *P*=.005). Therefore, employment should be controlled as a covariate in the hypothesized path model.

Path Model

After 3 months of the *Run4Love* intervention, 16 (5.3%) participants were lost to follow-ups, without their data of perceived stress and depressive symptoms at the 3- and 6-month assessments. The characteristics of the 16 participants were not significantly different from the remaining participants. The path model was estimated using data from the remaining 284 participants. The hypothesized model was examined, and there were 2 significant pathways: intervention→suicide and intervention→perceived stress→depressive symptoms→suicide. In this model, perceived stress did not appear to be significantly related to suicide (*beta*=−.05, *P*=.71), and the intervention did not affect depressive symptoms directly either (*beta*=−.06, *P*=.08). In addition, both employment and suicide at baseline were controlled as potential confounders in the hypothesized model, whereas only suicide showed statistical significance (*beta*=.44, *P*<.001). Consequently, employment as a covariate and the 2 nonsignificant pathways were removed in the final model.

The final path model showed good model fit (χ^2_4= 6.1, *P*=.19, CFI=0.99, TLI=0.97, RMSEA=0.04, and WRMR=0.55). Standardized regression coefficients for the final model are reported in Table 3 and Figure 3. Results indicated that the intervention reduced participants' perceived stress at the 3-month follow-up (*beta*=−.32, *P*<.001), which was positively associated with depressive symptoms at the 3-month follow-up (*beta*=.83, *P*<.001); reduced depressive symptoms consequently resulted in reduced suicide at 6 months (*beta*=.34, *P*<.001). The pathway from perceived stress to depressive symptoms had the strongest effect size in the mediation model (*beta*=.83, *P*<.001).

The direct, indirect, and total effects of the path model are summarized in Table 3. The direct effect of the intervention on suicide at the 6-month follow-up was significant (*beta*=−.18, *P*=.02). The indirect effect of the intervention on suicide via perceived stress and depressive symptoms was also significant (*beta*=−.32*.83*.34=−.09; *P*<.001). In summary, there was a partial mediating effect of perceived stress and depressive symptoms, accounting for 33% (−0.09/−0.27) of the total effect of the intervention on suicide.

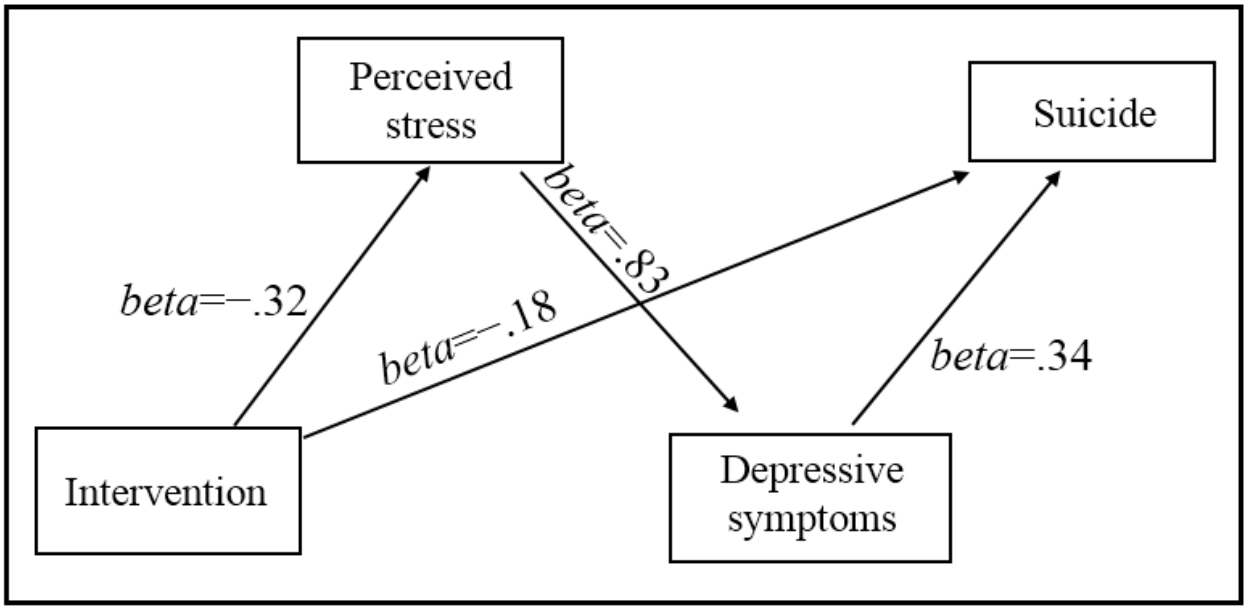
Table 3. Coefficients of the pathways in the final model (n=284).

Pathways	Coefficient (beta)	Standardized coefficient (beta)	95% CI	SE	*P* value
Intervention→ perceived stress[a]	−3.67	−.32	−5.03 to −2.31	0.69	<.001
Intervention→ suicide[b]	−.41	−.18	−0.76 to −0.07	0.18	.02
Perceived stress[a] → depressive symptoms[a]	1.44	.83	1.31 to 1.57	0.07	<.001
Depressive symptoms[a] → suicide[b]	.04	.34	0.02 to 0.06	0.01	<.001
Total effect	−.62	−.27	−0.96 to −0.28	0.17	<.001
Direct effect					
Intervention→ suicide[b]	−.41	−.18	−0.76 to −0.07	0.18	.02
Indirect effect					
Intervention→ perceived stress[a]→ depressive symptoms[a] → suicide[b]	−.21	−.09	−0.33 to −0.09	0.06	.001

[a]3-month follow-up.

[b]6-month follow-up.

Figure 3. Estimation of the final path model of intervention, perceived stress, depressive symptoms, and suicide in people living with HIV and depression.



Discussion

Principal Findings

This study was among the first efforts to explore the mechanisms of how an mHealth intervention reduced suicide in people living with HIV and depression, with a longitudinal design. Understanding the mechanisms is important to discern the processes between an intervention and mental health outcomes [40,41]. Better understanding of such mechanisms can facilitate targeted mHealth intervention to prevent suicide among people living with HIV and depression. Our data analyses revealed that the mHealth intervention had both direct and indirect effects on suicide, suggesting that the mHealth intervention had effectively reduced the participants' suicide. These findings demonstrated that suicide in people living with HIV and depression could be reduced through a WeChat-based CBSM intervention, such as *Run4Love*. The mechanisms of the indirect effects were mediated through reduced perceived stress and depressive symptoms, which consequently resulted in reduced suicide.

Implications of the Path Model

Consistent with the existing literature [15-18], perceived stress had a strong positive association with depressive symptoms, as evidenced by the large effect size shown in this study. This finding underscored the importance of reducing perceived stress to alleviate depressive symptoms in people living with HIV and depression. Many effective mHealth interventions that targeted depressive symptoms in people living with HIV have included stress management as a main component of the intervention, such as Web-based cognitive behavioral therapy [42].

The results also indicated the causal relationship between depressive symptoms at the 3-month follow-up and suicide at the 6-month follow-up. Consistent with previous studies conducted in the general population or psychiatric outpatients [5,6], the findings demonstrated that suicide might be effectively reduced through mitigating depressive symptoms. Furthermore, this study confirmed this relationship in the vulnerable population of people living with HIV and depression, who needed targeted intervention because of high rates of suicide. Some studies screened people living with HIV for depression before targeted intervention, and these demonstrated good efficacy [43-45]. Our data also suggested the importance of targeted suicide prevention intervention in people living with HIV and depression.

The direct effect of the mHealth intervention on suicide might also be explained by factors other than perceived stress and depressive symptoms, such as reduction of stigma and improvement of self-efficacy, positive coping, and social support. Further research is needed to better understand the roles and mechanisms of these potential factors on suicide.

Another finding of this study that needs to be noted is that people living with HIV and depression who committed suicide reported lower level of employment. Previous studies found that unemployment was a risk factor for suicide [46]. As a stressful event, unemployment could lead to financial distress and social isolation, which was associated with elevated stress and depression and, eventually, suicide [47,48]. Although the relationship was significant in bivariate analysis, it was no longer significant when other factors (perceived stress, depressive symptoms, and suicide at baseline) were controlled.

The data showed that the *Run4Love* mHealth intervention alone was effective, compared with the blank control group. Currently, the professional psychologists cannot meet the treatment demands of large number of people living with HIV and depression, especially in resources-limited settings, such as China. Therefore, an mHealth intervention can be an alternative resolution for people living with HIV and depression who cannot access psychotherapy. Alternatively, such an mHealth intervention can also be integrated with other traditional psychotherapies because of its easy access and adaptability. Future studies are needed to examine the effect of a stand-alone mHealth intervention compared with the one adjunct to psychotherapy.

This study suggested that public policies and coordinated efforts should be made to improve mental health outcomes, especially to reduce suicide among people living with HIV and depression. At the health care level, depression treatment and suicide prevention for people living with HIV are urgently needed. Although the guideline of HIV treatment has recommended integrating mental health screening into regular HIV care [49], such a practice was rarely implemented because of limited psychological recourses, perceived stigma of people living with HIV, and extra commute burden of face-to-face psychotherapy [42]. In middle- and low-income countries, such as China, mHealth interventions such as the WeChat-based *Run4Love* can reach more people in a cost-effective manner [13]. *Run4Love* platform also provides useful functions, such as automatic progress tracking and personalized feedback, which are necessary for an effective intervention. Our data underscore the importance of integrating mental health services into routine HIV health care by mobile apps, such as WeChat. At the community level, we call for more mHealth interventions, such as *Run4Love*, to reduce stress and improve mental health outcomes of people living with HIV and depression.

Limitations

Several limitations of this study should be noted. First, the participants were recruited from 1 hospital in a metropolitan area, and the majority of them were men. Therefore, the results might not be generalizable to all people living with HIV and depression in China, especially women or those in rural areas. Second, the measures in this study were self-reported, with potential recall or self-report biases. We did not include a standardized measurement of suicide, and we only had 2 items about suicidal ideation and suicidal attempt. Nevertheless, self-reported suicide assessment has shown a high level of agreement with a clinician-delivered face-to-face assessment, and this could serve as an efficient and reliable method to assess suicide outcome [50]. Third, like all modeling studies, assumptions were embedded in the current path model, for example, linear relationship among variables. Finally, the pathway model included only a limited number of covariates, and other potential factors that might affect mental health outcomes were not included.

Conclusions

In conclusion, this was the first study to examine the mechanisms of an mHealth intervention, *Run4Love*, in reducing suicide through both direct and indirect pathways, using longitudinal data from an RCT. Besides direct effect, the indirect effect was mediated by reduced perceived stress and depressive symptoms, which consequently resulted in reduced suicide. To reduce suicide in people living with HIV and depression, perceived stress and depressive symptoms are both key intervention targets. We call for targeted intervention to prevent suicide in people living with HIV and depression, especially mHealth programs that can reduce perceived stress.

Acknowledgments

This study was supported by a grant from the National Natural Science Foundation of China (Grant Number: 71573290) and China Medical Board (Grant Number: 17-271). The funder had no role in study design, data collection and analysis, or writing of the manuscript.

Authors' Contributions

YL analyzed the data and drafted the paper. YG and YAH contributed to funding obtaining, study design, and manuscript revision. MZ and CZ helped in study concept. JQ, ZX, HZ, and YZ contributed to clinical trial and data acquisition. WC, LL, and CL provided administrative, technical, and material support of the clinical trial.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-EHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 2277 KB - jmir_v21i11e14729_app1.pdf]

References

1. Bing EG, Burnam MA, Longshore D, Fleishman JA, Sherbourne CD, London AS, et al. Psychiatric disorders and drug use among human immunodeficiency virus-infected adults in the United States. Arch Gen Psychiatry 2001 Aug;58(8):721-728. [doi: 10.1001/archpsyc.58.8.721] [Medline: 11483137]
2. Wang T, Fu H, Kaminga AC, Li Z, Guo G, Chen L, et al. Prevalence of depression or depressive symptoms among people living with HIV/AIDS in China: a systematic review and meta-analysis. BMC Psychiatry 2018 May 31;18(1):160 [FREE Full text] [doi: 10.1186/s12888-018-1741-8] [Medline: 29855289]
3. UNAIDS. 2018. Global HIV & AIDS Statistics—2019 Fact Sheet. URL: http://www.unaids.org/en/resources/fact-sheet [accessed 2019-09-13]
4. Bengtson AM, Pence BW, Mimiaga MJ, Gaynes BN, Moore R, Christopoulos K, et al. Depressive symptoms and engagement in human immunodeficiency virus care following antiretroviral therapy initiation. Clin Infect Dis 2019 Jan 18;68(3):475-481 [FREE Full text] [doi: 10.1093/cid/ciy496] [Medline: 29901695]
5. Brown GK, Beck AT, Steer RA, Grisham JR. Risk factors for suicide in psychiatric outpatients: a 20-year prospective study. J Consult Clin Psychol 2000 Jun;68(3):371-377. [doi: 10.1037/0022-006X.68.3.371] [Medline: 10883553]
6. Zhang Y, Law CK, Yip PS. Psychological factors associated with the incidence and persistence of suicidal ideation. J Affect Disord 2011 Oct;133(3):584-590. [doi: 10.1016/j.jad.2011.05.003] [Medline: 21636133]
7. Badiee J, Moore DJ, Atkinson JH, Vaida F, Gerard M, Duarte NA, et al. Lifetime suicidal ideation and attempt are common among HIV+ individuals. J Affect Disord 2012 Feb;136(3):993-999 [FREE Full text] [doi: 10.1016/j.jad.2011.06.044] [Medline: 21784531]
8. Zeng C, Li L, Hong YA, Zhang H, Babbitt AW, Liu C, et al. A structural equation model of perceived and internalized stigma, depression, and suicidal status among people living with HIV/AIDS. BMC Public Health 2018 Jan 15;18(1):138 [FREE Full text] [doi: 10.1186/s12889-018-5053-1] [Medline: 29334959]
9. Cooper V, Clatworthy J, Whetham J, Consortium E. mHealth interventions to support self-management in HIV: a systematic review. Open AIDS J 2017;11:119-132 [FREE Full text] [doi: 10.2174/1874613601711010119] [Medline: 29290888]
10. Cao B, Gupta S, Wang J, Hightow-Weidman LB, Muessig KE, Tang W, et al. Social media interventions to promote HIV testing, linkage, adherence, and retention: systematic review and meta-analysis. J Med Internet Res 2017 Nov 24;19(11):e394 [FREE Full text] [doi: 10.2196/jmir.7997] [Medline: 29175811]
11. Kempf M, Huang C, Savage R, Safren SA. Technology-delivered mental health interventions for people living with HIV/AIDS(PLWHA): a review of recent advances. Curr HIV/AIDS Rep 2015 Dec;12(4):472-480 [FREE Full text] [doi: 10.1007/s11904-015-0292-6] [Medline: 26452644]
12. Swendeman D, Ramanathan N, Baetscher L, Medich M, Scheffler A, Comulada WS, et al. Smartphone self-monitoring to support self-management among people living with HIV: perceived benefits and theory of change from a mixed-methods randomized pilot study. J Acquir Immune Defic Syndr 2015 May 1;69 Suppl 1:S80-S91 [FREE Full text] [doi: 10.1097/QAI.0000000000000570] [Medline: 25867783]
13. van Luenen S, Garnefski N, Spinhoven P, Kraaij V. Guided internet-based intervention for people with HIV and depressive symptoms: a randomised controlled trial in the Netherlands. Lancet HIV 2018 Sep;5(9):e488-e497. [doi: 10.1016/S2352-3018(18)30133-4] [Medline: 30135045]
14. Guo Y, Hong YA, Qiao J, Xu Z, Zhang H, Zeng C, et al. Run4Love, a mHealth (WeChat-based) intervention to improve mental health of people living with HIV: a randomized controlled trial protocol. BMC Public Health 2018 Jun 26;18(1):793 [FREE Full text] [doi: 10.1186/s12889-018-5693-1] [Medline: 29940921]
15. Metalsky GI, Joiner TE. Vulnerability to depressive symptomatology: a prospective test of the diathesis-stress and causal mediation components of the hopelessness theory of depression. J Pers Soc Psychol 1992 Oct;63(4):667-675. [doi: 10.1037//0022-3514.63.4.667] [Medline: 1447690]
16. Brown TA, Rosellini AJ. The direct and interactive effects of neuroticism and life stress on the severity and longitudinal course of depressive symptoms. J Abnorm Psychol 2011 Nov;120(4):844-856 [FREE Full text] [doi: 10.1037/a0023035] [Medline: 21381799]

17. Felsten G. Stress reactivity and vulnerability to depressed mood in college students. Pers Indiv Differ 2004;36(4):789-800. [doi: 10.1016/S0191-8869(03)00152-1]
18. Caspi A, Sugden K, Moffitt TE, Taylor A, Craig IW, Harrington H, et al. Influence of life stress on depression: moderation by a polymorphism in the 5-HTT gene. Science 2003 Jul 18;301(5631):386-389 [FREE Full text] [doi: 10.1126/science.1083968] [Medline: 12869766]
19. Smith SS, Smith Carter J, Karczewski S, Pivarunas B, Suffoletto S, Munin A. Mediating effects of stress, weight-related issues, and depression on suicidality in college students. J Am Coll Health 2015;63(1):1-12. [doi: 10.1080/07448481.2014.960420] [Medline: 25222880]
20. Anastasiades MH, Kapoor S, Wootten J, Lamis DA. Perceived stress, depressive symptoms, and suicidal ideation in undergraduate women with varying levels of mindfulness. Arch Womens Ment Health 2017 Feb;20(1):129-138. [doi: 10.1007/s00737-016-0686-5] [Medline: 27778150]
21. Gill SK, Muñoz RF, Leykin Y. The influence of perceived stress and depression on suicide-related beliefs in caucasian and Indian adults. Crisis 2018 Mar;39(2):127-136. [doi: 10.1027/0227-5910/a000492] [Medline: 29216756]
22. Guangdong Provincial Center for Disease Control and Prevention. Guangdong Provincial Center for Disease Control and Prevention. 2017. AIDS Prevention and Treatment in 2017 of Guangdong Province (in China). URL: http://cdcp.gd.gov.cn/hhff/inbfo/content/post_1107865.html [accessed 2019-09-13] [WebCite Cache ID 78BzomWlG]
23. Antoni M, Ironson G, Schneiderman N. Cognitive-behavioral Stress Management For Individuals Living With HIV (Treatments That Work). Oxford: Oxford University Press; 2007.
24. Cruess DG, Antoni MH, Schneiderman N, Ironson G, McCabe P, Fernandez JB, et al. Cognitive-behavioral stress management increases free testosterone and decreases psychological distress in HIV-seropositive men. Health Psychol 2000 Jan;19(1):12-20. [doi: 10.1037/0278-6133.19.1.12] [Medline: 10711583]
25. McGregor BA, Dolan ED, Murphy KM, Sannes TS, Highland KB, Albano DL, et al. Cognitive behavioral stress management for healthy women at risk for breast cancer: a novel application of a proven intervention. Ann Behav Med 2015 Dec;49(6):873-884 [FREE Full text] [doi: 10.1007/s12160-015-9726-z] [Medline: 26290001]
26. Gaab J, Blättler N, Menzi T, Pabst B, Stoyer S, Ehlert U. Randomized controlled evaluation of the effects of cognitive-behavioral stress management on cortisol responses to acute stress in healthy subjects. Psychoneuroendocrinology 2003 Aug;28(6):767-779. [doi: 10.1016/s0306-4530(02)00069-0] [Medline: 12812863]
27. Tencent. 2018. Tencent Announces 2018 Second Quarter and Interim Results. URL: https://www.tencent.com/en-us/articles/8003521534381984.pdf [accessed 2019-05-07] [WebCite Cache ID 78BxzHm7v]
28. Radloff LS. The CES-D Scale:A self-report depression scale for research in the general population. Appl Psychol Meas 2016 Jul 26;1(3):385-401. [doi: 10.1177/014662167700100306]
29. Wang M, Armour C, Wu Y, Ren F, Zhu X, Yao S. Factor structure of the CES-D and measurement invariance across gender in mainland Chinese adolescents. J Clin Psychol 2013 Sep;69(9):966-979. [doi: 10.1002/jclp.21978] [Medline: 23775279]
30. Zhang Y, Ting RZ, Lam MH, Lam S, Yeung RO, Nan H, et al. Measuring depression with CES-D in Chinese patients with type 2 diabetes: the validity and its comparison to PHQ-9. BMC Psychiatry 2015 Aug 18;15:198 [FREE Full text] [doi: 10.1186/s12888-015-0580-0] [Medline: 26281832]
31. Zich JM, Attkisson CC, Greenfield TK. Screening for depression in primary care clinics: the CES-D and the BDI. Int J Psychiatry Med 1990;20(3):259-277. [doi: 10.2190/LYKR-7VHP-YJEM-MKM2] [Medline: 2265888]
32. Lu W, Bian Q, Wang W, Wu X, Wang Z, Zhao M. Chinese version of the Perceived Stress Scale-10: A psychometric study in Chinese university students. PLoS One 2017;12(12):e0189543 [FREE Full text] [doi: 10.1371/journal.pone.0189543] [Medline: 29252989]
33. Leung DY, Lam T, Chan SS. Three versions of Perceived Stress Scale: validation in a sample of Chinese cardiac patients who smoke. BMC Public Health 2010 Aug 25;10:513 [FREE Full text] [doi: 10.1186/1471-2458-10-513] [Medline: 20735860]
34. Wang Z, Chen J, Boyd JE, Zhang H, Jia X, Qiu J, et al. Psychometric properties of the Chinese version of the Perceived Stress Scale in policewomen. PLoS One 2011;6(12):e28610 [FREE Full text] [doi: 10.1371/journal.pone.0028610] [Medline: 22164311]
35. Ng SM. Validation of the 10-item Chinese perceived stress scale in elderly service workers: one-factor versus two-factor structure. BMC Psychol 2013;1(1):9 [FREE Full text] [doi: 10.1186/2050-7283-1-9] [Medline: 25566361]
36. Olpin M, Hesson M. Stress Management For Life: A Research-based Experiential Approach. Boston: Cengage Learning; 2015.
37. Hu L, Bentler PM. Fit indices in covariance structure modeling: sensitivity to underparameterized model misspecification. Psychol Methods 1998;3(4):424-453. [doi: 10.1037//1082-989X.3.4.424]
38. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct Equ Modeling 1999 Jan;6(1):1-55. [doi: 10.1080/10705519909540118]
39. Kline RB. Latent variable path analysis in clinical research: a beginner's tour guide. J Clin Psychol 1991 Jul;47(4):471-484. [doi: 10.1002/1097-4679(199107)47:4<471::aid-jclp2270470402>3.0.co;2-o] [Medline: 1939690]

40. Silverstein M, Cabral H, Hegel M, Diaz-Linhart Y, Beardslee W, Kistin CJ, et al. Problem-solving education to prevent depression among low-income mothers: a path mediation analysis in a randomized clinical trial. JAMA Netw Open 2018 Jun 1;1(2):e180334 [FREE Full text] [doi: 10.1001/jamanetworkopen.2018.0334] [Medline: 30646083]
41. Carrico AW, Ironson G, Antoni MH, Lechner SC, Durán RE, Kumar M, et al. A path model of the effects of spirituality on depressive symptoms and 24-h urinary-free cortisol in HIV-positive persons. J Psychosom Res 2006 Jul;61(1):51-58. [doi: 10.1016/j.jpsychores.2006.04.005] [Medline: 16813845]
42. Karyotaki E. Internet-based interventions for people with HIV and depression. Lancet HIV 2018 Sep;5(9):e474-e475. [doi: 10.1016/S2352-3018(18)30172-3] [Medline: 30135044]
43. Kelly JA, Murphy DA, Bahr GR, Kalichman SC, Morgan MG, Stevenson LY, et al. Outcome of cognitive-behavioral and support group brief therapies for depressed, HIV-infected persons. Am J Psychiatry 1993 Nov;150(11):1679-1686. [doi: 10.1176/ajp.150.11.1679] [Medline: 8214177]
44. Lee MR, Cohen L, Hadley SW, Goodwin FK. Cognitive-behavioral group therapy with medication for depressed gay men with AIDS or symptomatic HIV infection. Psychiatr Serv 1999 Jul;50(7):948-952. [doi: 10.1176/ps.50.7.948] [Medline: 10402618]
45. Kraaij V, van Emmerik A, Garnefski N, Schroevers MJ, Lo-Fo-Wong D, van Empelen P, et al. Effects of a cognitive behavioral self-help program and a computerized structured writing intervention on depressed mood for HIV-infected people: a pilot randomized controlled trial. Patient Educ Couns 2010 Aug;80(2):200-204. [doi: 10.1016/j.pec.2009.08.014] [Medline: 19781889]
46. Milner A, Page A, LaMontagne AD. Cause and effect in studies on unemployment, mental health and suicide: a meta-analytic and conceptual review. Psychol Med 2014 Apr;44(5):909-917. [doi: 10.1017/S0033291713001621] [Medline: 23834819]
47. Yur'yev A, Värnik A, Värnik P, Sisask M, Leppik L. Employment status influences suicide mortality in Europe. Int J Soc Psychiatry 2012 Jan;58(1):62-68. [doi: 10.1177/0020764010387059] [Medline: 21088037]
48. Schneider B, Grebner K, Schnabel A, Hampel H, Georgi K, Seidler A. Impact of employment status and work-related factors on risk of completed suicide. A case-control psychological autopsy study. Psychiatry Res 2011 Dec 30;190(2-3):265-270. [doi: 10.1016/j.psychres.2011.07.037] [Medline: 21890214]
49. World Health Organization. Consolidated Guidelines On The Use Of Antiretroviral Drugs For Treating And Preventing HIV Infection: Recommendations For A Public Health Approach. Geneva: WHO press; 2016.
50. Safer DJ. Self-reported suicide attempts by adolescents. Ann Clin Psychiatry 1997 Dec;9(4):263-269. [Medline: 9511952]

Abbreviations

CBSM: cognitive-behavioral stress management
CES-D: Center for Epidemiologic Studies-Depression
CFI: Comparative Fit Index
IQR: interquartile range
mHealth: mobile health
PSS: Perceived Stress Scale
RCT: randomized controlled trial
RMSEA: root mean square error of approximation
TLI: Tucker-Lewis Index
WRMR: weighted root mean square residual

Edited by G Eysenbach; submitted 18.05.19; peer-reviewed by R Melia, L Elopre; comments to author 03.08.19; revised version received 25.08.19; accepted 31.08.19; published 27.11.19.

Please cite as:
Li Y, Guo Y, Hong YA, Zhu M, Zeng C, Qiao J, Xu Z, Zhang H, Zeng Y, Cai W, Li L, Liu C
Mechanisms and Effects of a WeChat-Based Intervention on Suicide Among People Living With HIV and Depression: Path Model Analysis of a Randomized Controlled Trial
J Med Internet Res 2019;21(11):e14729
URL: https://www.jmir.org/2019/11/e14729
doi:10.2196/14729
PMID:31774411

Review

Relating Instructional Design Components to the Effectiveness of Internet-Based Mindfulness Interventions: A Critical Interpretive Synthesis

Marie Lippmann[1*], MSc, PhD; Helena Laudel[2*], BSc, MSc; Marlene Heinzle[3], BSc, MA; Susanne Narciss[2], Dr Phil

[1]Department of Psychology, California State University, Chico, CA, United States

[2]Psychology of Learning and Instruction, Faculty of Psychology - School of Science, Technische Universität Dresden, Dresden, Germany

[3]Department of Media, Cognition and Communication, Faculty of Humanities, University of Copenhagen, Copenhagen, Denmark

[*]these authors contributed equally

Corresponding Author:
Marie Lippmann, MSc, PhD
Department of Psychology
California State University
400 West First Street
Chico, CA, 95929
United States
Phone: 1 530 898 5281
Email: mlippmann@csuchico.edu

Abstract

Background: Internet-based mindfulness interventions are a promising approach to address challenges in the dissemination and implementation of mindfulness interventions, but it is unclear how the instructional design components of such interventions are associated with intervention effectiveness.

Objective: The objective of this study was to identify the instructional design components of the internet-based mindfulness interventions and provide a framework for the classification of those components relative to the intervention effectiveness.

Methods: The critical interpretive synthesis method was applied. In phase 1, a strategic literature review was conducted to generate hypotheses for the relationship between the effectiveness of internet-based mindfulness interventions and the instructional design components of those interventions. In phase 2, the literature review was extended to systematically explore and revise the hypotheses from phase 1.

Results: A total of 18 studies were identified in phase 1; 14 additional studies were identified in phase 2. Of the 32 internet-based mindfulness interventions, 18 were classified as more effective, 11 as less effective, and only 3 as ineffective. The effectiveness of the interventions increased with the level of support provided by the instructional design components. The main difference between effective and ineffective interventions was the presence of just-in-time information in the form of reminders. More effective interventions included more supportive information (scores: 1.91 in phases 1 and 2) than less effective interventions (scores: 1.00 in phase 1 and 1.80 in phase 2), more part-task practice (scores: 1.18 in phase 1 and 1.60 in phase 2) than less effective interventions (scores: 0.33 in phase 1 and 1.40 in phase 2), and provided more just-in-time information (scores: 1.35 in phase 1 and 1.67 in phase 2) than less effective interventions (scores: 0.83 in phase 1 and 1.60 in phase 2). The average duration of more effective, less effective, and ineffective interventions differed for the studies of phase 1, with more effective interventions taking up more time (7.45 weeks) than less effective (4.58 weeks) or ineffective interventions (3 weeks). However, this difference did not extend to the studies of phase 2, with comparable average durations of effective (5.86 weeks), less effective (5.6 weeks), and ineffective (7 weeks) interventions.

Conclusions: Our results suggest that to be effective, internet-based mindfulness interventions must contain 4 instructional design components: formal learning tasks, supportive information, part-task practice, and just-in-time information. The effectiveness of the interventions increases with the level of support provided by each of these instructional design components.

(J Med Internet Res 2019;21(11):e12497) doi:10.2196/12497

KEYWORDS

mindfulness; internet; instructional design

Introduction

Background

Many medical conditions are accompanied by experiences of discomfort, worry, rumination, and anxiety [1,2]. The practice of mindfulness helps individuals who suffer from medical and psychological conditions by decreasing the perceived effects of their symptoms and increasing psychological well-being [3,4]. Central to the concept of mindfulness is the idea of cultivating a nonjudgmental and accepting awareness of present experiences (ie, thoughts, feelings, and bodily sensations) as they arise [5]. As mindfulness is a complex skill that requires learning and practice, mindfulness-based interventions have been developed and implemented to assist individuals in mastering this skill [6,7]. Mindfulness-based interventions are typically administered in person but face challenges in terms of their dissemination and implementation. Internet-based mindfulness interventions are a promising new approach to address these challenges [8]. Several recent reviews provide insights into the effectiveness of internet-based mindfulness interventions for a variety of outcome measures [9-15]. However, no research has yet investigated how the design of those interventions is associated with intervention effectiveness. This study closes this gap in the literature by identifying instructional design components of internet-based mindfulness interventions and providing a framework for the classification of those components, relative to the intervention effectiveness.

Internet-Based Mindfulness Interventions

Internet-based mindfulness interventions, as a subcategory of internet-based health interventions, have the potential to reach a large number of potential users, extend intervention accessibility to individuals with economic and transportation restrictions, increase intervention convenience through greater flexibility in use and application, avoid social stigma of therapeutic settings, and increase cost-effectiveness for both providers and clients [16-18]. In the past 5 years, 7 reviews have systematically investigated the effectiveness of internet-based mindfulness interventions.

Spijkerman and Bohlmeijer [9] examined 15 randomized controlled trials, comparing internet-based mindfulness interventions with control conditions. They found the internet-based mindfulness interventions to have significant small-to-moderate effects on mental health.

Fish et al [10] reviewed 10 technology-based mindfulness interventions aimed at clinical outcomes of mental health (stress, depression, and anxiety) and found that 8 studies produced significant effects but with varying effect sizes. The authors point out that they found it difficult to draw conclusions about intervention effectiveness relative to design components of the interventions, such as construction, length, and delivery, and explicitly call for further research to investigate this issue [10].

Toivonen et al [11] reviewed 16 internet-based mindfulness interventions aimed at physiological symptoms (eg, cancer, chronic pain or fibromyalgia, irritable bowel syndrome, epilepsy, heart disease, tinnitus, and acquired brain injury). They found that the majority of the studies reported positive effects of the internet-based mindfulness interventions compared with traditional treatment on a multitude of outcomes, including pain acceptance, coping mechanisms, and symptoms of depression [11]. The authors found mixed results when comparing internet-based mindfulness interventions to active control groups receiving, for example, cognitive behavioral therapy [11].

Heber et al [12] reviewed the effectiveness of internet- and computer-based stress management interventions in a meta-analysis including 26 comparisons. The authors found large effect sizes for the investigated interventions, relative to control groups, in terms of stress reduction; small effects were obtained for depression [12]. Subgroup analyses revealed that guided interventions were more effective than unguided interventions, and the authors found differences in intervention effectiveness based on the design characteristics of duration and intervention content [12], thereby highlighting the need for more research on the design of internet-based interventions and potential relationships between design and intervention effectiveness.

Lyzwinski et al [13] reviewed 21 internet-based mindfulness interventions for stress, maladaptive weight-related behaviors, and weight loss. They found that most interventions were effective for stress reduction. Conclusions about intervention effectiveness for weight-related behaviors could not be drawn because not enough studies with weight-related outcomes were identified [13].

Mikolasek et al [14] reviewed 17 empirical studies on internet-based mindfulness or relaxation interventions for medical conditions (eg, irritable bowel syndrome, cancer, chronic pain, surgery, and hypertension). This review found that the internet-based mindfulness or relaxation interventions were mostly effective, with varying effect sizes, but it found no effects for stress [14]. In the discussion of their findings, the authors point to differences in intervention design, such as intervention dose and regularity, as potential sources for the heterogeneity in intervention effectiveness [14].

Finally, Sevilla-Llewellyn-Jones et al [15] reviewed 12 internet-based mindfulness interventions for mental health in clinical populations and found that the internet-based mindfulness interventions were effective in reducing depression and anxiety while enhancing the quality of life and mindfulness skills, particularly in individuals with clinical anxiety. The authors point to challenges in the interpretation of the results based on the heterogeneity of the interventions and their components [15], providing further incentive to investigate the design of internet-based mindfulness interventions.

Overall, the reviews show heterogeneous, but predominantly encouraging, results in support of the effectiveness of internet-based mindfulness interventions aimed at a variety of mental and physical health conditions [9-15]. The authors of

the majority of those reviews point out that the extent to which such findings can be generalized is limited by the large variety in components of internet-based interventions, including differences in content, scheduling, guidance, and support [10,12,14,15]. As a result, it is unclear which design components are associated with intervention effectiveness, and more research is needed to investigate the design of internet-based mindfulness interventions relative to their effectiveness.

Relevant Components of Internet-Based Mindfulness Interventions: Instructional Design Perspective

Employing instructional design process models increases learning outcomes across a variety of contexts [19]. The need for instructional design in developing internet-based interventions is based on the premise that technology components are more likely to have positive effects on learning processes and outcomes when intervention designers take a learner-centered and need-based approach [19]. As mindfulness can be viewed as a complex skill demanding extensive amounts of practice to be learned and mastered [5], instructional design models provide powerful tools to identify relevant design components. An instructional design model that is particularly suitable to apply to complex learning processes, such as establishing and mastering mindfulness, is the 4-component instructional design (4C/ID) model [20]. The 4C/ID model comprises 4 core components.

The first component comprises *learning tasks* (LTs) that are authentic whole-task experiences. In the context of internet-based mindfulness interventions, LTs are represented by formal mindfulness exercises, typically guided meditations in the format of audio files. LTs are further specified in terms of their content and scheduling and whether they are tailored to specific conditions or target specific populations [21,22]. All of these aspects of LTs are relevant to intervention effects [21-23] and are, therefore, considered in this review.

The second component in the 4C/ID model refers to *supportive information* (SI) assisting learners in the acquisition and performance of nonrecurrent aspects of the LT to help establish correct mental models and appropriate cognitive strategies. In the context of internet-based mindfulness interventions, SI is represented by reflection exercises, psychoeducative information, and peer support in forums or chat rooms. SI seems to facilitate intervention effects for internet-based health interventions, but the effects are difficult to gauge because the SI is typically presented in addition to the main intervention [24].

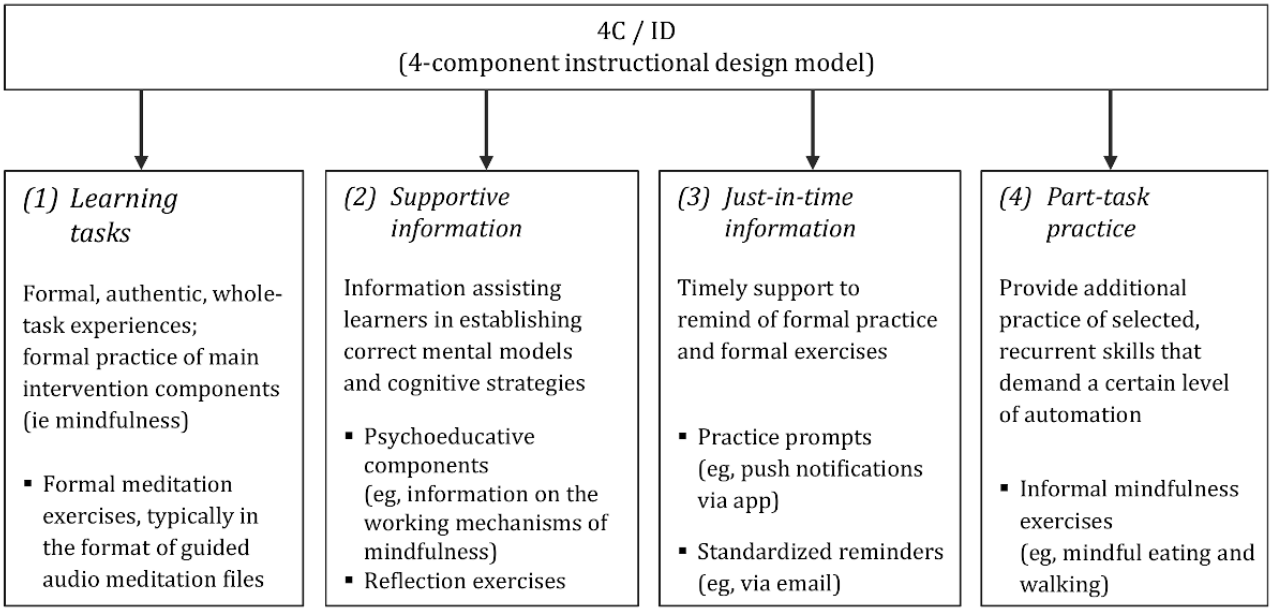
The third component in the 4C/ID model refers to *just-in-time information* (JIT) that concerns information supporting the performance of recurrent aspects of the LT.

In terms of self-help programs, Cavanagh et al [25] found larger effects for programs with guiding prompts than for unguided programs. In internet-based mindfulness interventions, JIT is represented by prompts and reminders encouraging continuous, regular practice.

The fourth component in the 4C/ID model comprises *part-task practice* (PTP) that refers to the additional practice of selected recurrent skills that demand a certain level of automation. In internet-based mindfulness interventions, PTP is represented by informal practice exercises aimed at practicing the established mindfulness skills during everyday activities, such as mindful eating or walking.

In face-to-face interventions, both formal and informal practices have been recognized to play a key role in the development of mindfulness [3] and are, therefore, considered relevant to this review. For an overview of the components of internet-based mindfulness interventions mapped onto the components of the 4C/ID model, see Figure 1.

Figure 1. Four-component instructional design model.



Research Questions

With reference to the 4C/ID model, we applied the critical interpretive synthesis (CIS) method [26] to address the following 2 research questions:

1. *Which instructional design components can be identified in existing internet-based mindfulness interventions?*
2. *How can these design components be classified relative to the intervention effectiveness?*

Methods

The Critical Interpretive Synthesis Method

A typical approach to answering research questions from the existing literature is the Cochrane-style systematic review [26]. The CIS method provides an alternative to systematic reviews whenever the literature does not provide a sufficient foundation for a meta-analysis. As thus far, no empirical studies have investigated the instructional design of internet-based mindfulness interventions, we utilized strategic elements of systematic reviews but implemented those within the more flexible CIS method [21]. In contrast to systematic reviews, CIS does not rely on exhaustive literature searches, rigid inclusion criteria, and quality assessments but employs techniques from qualitative research, such as diversity sampling, to generate hypotheses and systematically explore those in an iterative and dynamic review process [26]. To address our research questions, we implemented CIS in 2 phases.

Phase 1: Diversity Sampling and Generation of Hypotheses

Aim

The aim in phase 1 was to obtain a diverse sample of the existing literature on internet-based mindfulness interventions to identify instructional design components and generate a framework for classifying the effectiveness of the interventions relative to their design components.

Inclusion and Exclusion Criteria

With regard to population, we applied no restriction criteria because the transdiagnostic applicability of mindfulness is likely to result in a great variability of outcomes for varying populations, and we were interested in obtaining a diverse sample of the literature. With regard to the inclusion criteria for the intervention, we defined that interventions had to be delivered through the internet (ie, via a website or a mobile phone app) and contain formal mindfulness exercises as their main component to ensure the comparability of LTs across studies. In consequence, studies employing multicomponent interventions such as the Acceptance and Commitment Therapy [26] were not considered. Furthermore, interventions had to be delivered asynchronously, excluding interventions with live delivery, for example, via videoconferencing, to ensure the comparability of guidance and support components across studies. Only studies with control groups (typically waitlist controls) were considered. The intervention outcomes had to be indicative of mental or physical health. Relevant empirical studies had to have been published in an international peer-reviewed journal in English language. As both mindfulness- and internet-based interventions have predominantly emerged within the past two decades, no time restrictions were applied.

Search Strategy

A literature search was conducted in the databases PsycINFO, PsycARTICLES, PubMed, and Web of Science between February 2016 and October 2018.

After conducting a set of preliminary searches to identify the most accurate key words, the following search terms were included in all 4 databases (*mindfulness OR *mindful OR *meditation) AND (*internet OR *web OR *online OR *smartphone OR *app OR *mobile). The titles and abstracts of all search results were first screened for relevance. After removing duplicates of the publications identified as relevant, unavailable results were requested from the authors. On the basis of the results identified as dissertations, an additional author search was conducted to determine if the reported trials had been published in the meantime. All remaining studies were then subjected to a full-text screening. In the last step, the references of the included studies were screened to identify relevant research not covered by the database searches.

Quality Assessment in Terms of Risk Bias

Risk bias was assessed with the guidelines developed by the Cochrane Back Review Group [27] that had been successfully employed in a previous review on mindfulness-based interventions [28]. Assessment criteria focused on whether (1) methods of randomization were reported, (2) intervention outcomes were assessed with standardized measures, (3) a follow-up assessment was performed, (4) analyses included an intention-to-treat analysis, (5) sample characteristics were reported, (6) characteristics of withdrawals and dropouts were reported, and (7) studies contained detailed intervention descriptions. For each of the 7 criteria, 2 points were awarded if the criterion was reported and adhered to, 1 point if the criterion was reported but not adhered to, and 0 points if the criterion was not reported. The sum of the awarded points serves as an indicator of study quality, with 0 to 7 points indicating low, 8 to 11 indicating moderate, and 12 to 14 indicating high quality.

Review Strategy

For each included study, general information including authors, publication year, and country was recorded. The quality of the included studies was assessed in terms of risk bias. Sample characteristics including gender, age, and medical indications were retrieved. Outcome measures and characteristics of the control group were recorded. Group differences were assessed in terms of between- and within-group effects. The ranges of the effect sizes for the main outcome measures were recorded whenever effect sizes were reported in the original studies. Whenever effect sizes were not reported, we recorded reported *P* values instead. The effectiveness of the interventions was assessed with the criteria for defining intervention effectiveness [21]. The operationalization of those criteria is shown in Table 1.

The instructional design components of each intervention were identified and mapped onto the 4 components of the 4C/ID

model (ie, LTs, SI, JIT, and PTP). In addition, information on the duration and scheduling of the interventions were recorded, and reports of adherence and acceptance were included whenever they were reported in the original studies.

Table 1. Criteria for defining intervention effectiveness according to Morrison et al (2012).

Intervention code	Criteria
More effective	• The intervention led to improvement on majority of outcomes measures. • The intervention was at least as effective as comparison groups. • The intervention was more effective than waiting list or no intervention control groups.
Less effective	• The intervention led to improvement on minority of outcomes measures. • The intervention was not necessarily as effective as comparison groups. • The intervention was more effective than waiting list or no intervention control groups.
Ineffective	• The intervention did not lead to improvement on any of the outcome measures. • The intervention was no more effective than waiting list or no intervention control groups.

Generation of Hypotheses for Phase 2

To systematically evaluate the instructional design components of the interventions relative to the intervention effectiveness, we constructed and implemented the following scoring system: each intervention received points in the range from 0 to 2 for each of the 4C/ID components. The operationalization of those points relative to the 4C/ID components is reported in Table 2.

Each intervention was scored according to this system. In a next step, average scores were computed for each of the 4C/ID components and mapped onto intervention effectiveness. These rating processes were conducted by two independent raters, and interrater reliability was computed.

On the basis of the review of this first set of studies, we systematically generated hypotheses regarding the association between instructional design components of internet-based mindfulness interventions and intervention effectiveness. According to the CIS procedure [26], a second literature search (representative sampling) was then conducted to explore whether the hypotheses from phase 1 were generalizable and consistent.

Table 2. Intervention ratings for 4-component instructional design components (duration was 1 point per week, and in case of varying data count, the duration was longest).

Score (points)	Learning task	Supportive information	Part-task practice	Just-in-time information
0	Not existent	Not existent	Not existent	Not existent
1	Existent but not described; formal exercises implemented less than twice per week; formal exercises stable in content	Existent but not described; educational material provided only once; optional contact in case of questions or problems	Existent but not described; 1 informal exercise on a single distinct topic; exercises only once per week or less	Existent but not described; reminders once per week or less; reminders only when adherence was absent
2	Described formal exercises with varying content; implemented at least twice per week	Continuously accessible educational, supportive material; reflection exercises (eg, as diary or log writing)	Several unstructured informal exercises across a variety of topics; implemented at least twice per week	1 reminder ahead of each scheduled practice; adjustable reminders; prompts with monitoring information

Phase 2: Representative Sampling

Aim

The aim of phase 2 was to explore the extent to which the hypotheses from phase 1 were consistent and generalizable across an additional set of empirical studies.

Inclusion and Exclusion Criteria

Inclusion and exclusion criteria were identical to those in phase 1.

Search Strategy

To identify a representative sample of studies that may include more recent research, a systematic literature search was conducted to identify systematic reviews on the topic of internet-based mindfulness interventions published over the last 4 years. The studies reported in those reviews were identified, inclusion and exclusion criteria were applied, and the remaining studies were compared with the studies from the search in phase 1. This comparison revealed that all studies from the reviews matching our inclusion criteria were already represented in our search in phase 1. Therefore, a hand search in JMIR-relevant journals of the past 3 years was conducted, and the databases PubMed, PsycARTICLES, and PsycINFO were searched again for more recent studies in the years 2017 to 2019. For this hand search, the search string from phase 1 was used, titles and abstracts were scanned for relevance, duplicates were removed, and the remaining articles were subjected to a full-text screening.

Review Strategy

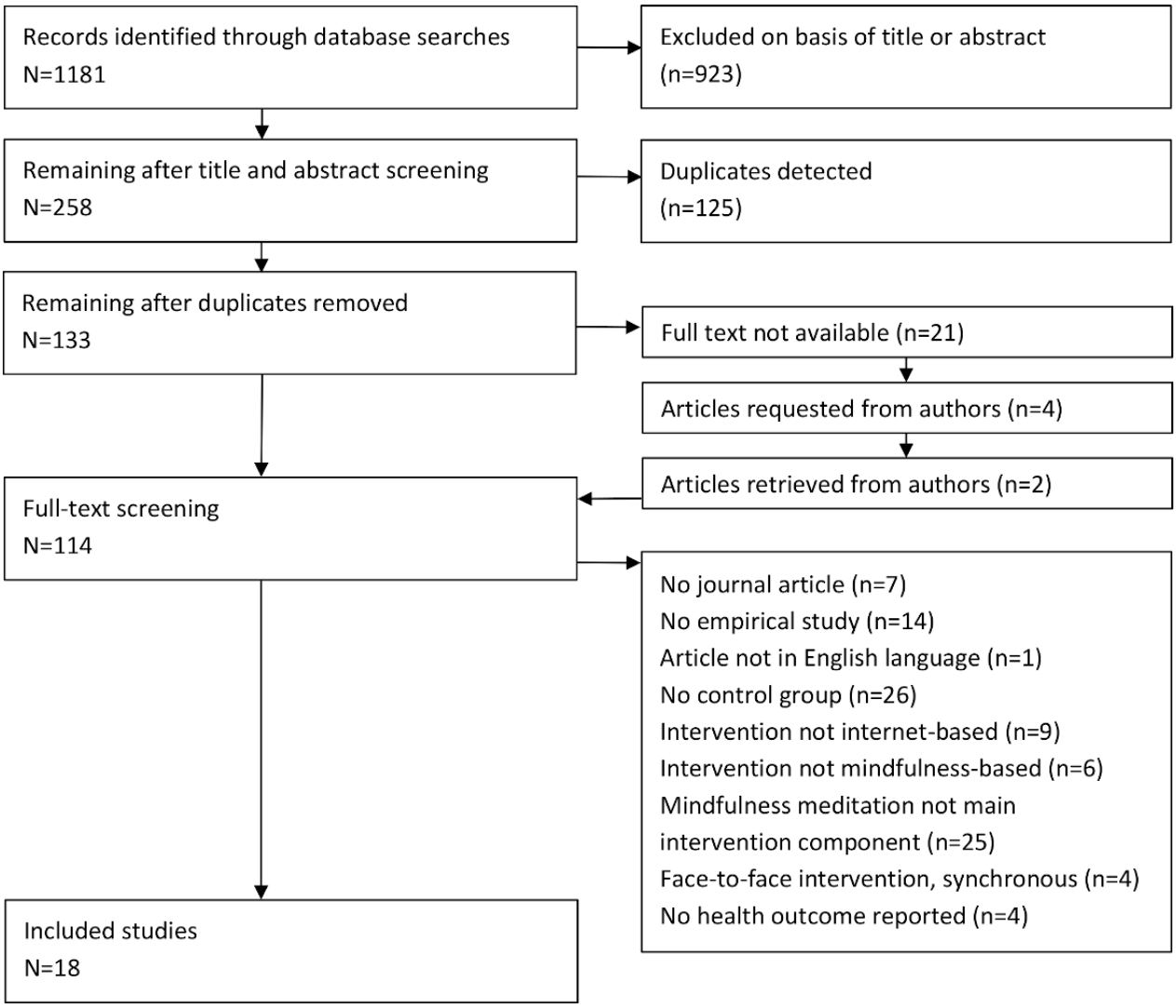
The review strategy from phase 1 was applied again, with the goal of identifying any instructional design components that might have not occurred in phase 1. Then, the associations between instructional design components and intervention effectiveness were examined.

Results

Phase 1: Diversity Sampling and Generation of Hypotheses

The search process from the keyword search to the final study selection is visualized in Figure 2. The systematic literature search across the 4 databases revealed 1181 results. In the title and abstract screening, 258 search results were identified as relevant, of which 125 were identified as duplicates. Of the remaining 133 studies, 112 were accessible. Of the nonaccessible 21 studies, 7 were identified as dissertations and conference papers, 3 did not employ control groups, and 2 were nonempirical. Of the remaining 5 nonaccessible studies, 1 study did not include author information. The remaining 4 nonaccessible studies were requested directly from the authors via email, with a return of 2 studies that were then added to the pool of studies for a full-text screening. This procedure resulted in 114 studies that were subjected to a full-text screening with regard to the defined inclusion and exclusion criteria. Of these 114 studies, 18 studies matched the inclusion criteria and were, therefore, included in this systematic review [29-46].

Figure 2. The study selection process in phase 1.



General Description and Intervention Effectiveness

The general descriptions of the characteristics of the 18 studies included in phase 1 are reported in detail for each study (authors, year of publication, study design, follow-up measures, sample size, age, and indication) in Multimedia Appendix 1. The quality scores in terms of risk bias are also reported in Multimedia Appendix 1. In summary, 3 studies achieved a quality score of 14, 9 a score of 13, 3 a score of 12, 1 a score of 10, and 2 a score of 9. The majority of interventions (n=12) were aimed at psychological symptoms such as stress, anxiety, depression, and general well-being. A total of 4 interventions were aimed at physiological symptoms (eg, fibromyalgia, chronic pain, and heart disease). A total of 2 interventions were aimed at work-related issues (eg, work-life balance and work-related well-being).

A detailed overview of intervention effectiveness (ie, outcome measures, control groups, within- and between-group effects, and computed effectiveness ratings or ERs) is reported in Multimedia Appendix 2. Upon applying the criteria for defining intervention effectiveness [21], 11 studies were classified as more effective, 6 as less effective, and 1 as ineffective.

Intervention Design and Effectiveness

The intervention design, duration and scheduling, and adherence and acceptance of the interventions are reported in detail in Multimedia Appendix 3. For an overview of intervention design relative to its effectiveness, ERs are reported again in this overview.

Intervention Ratings for 4-Component Instructional Design Components

Tables 3 and 4 report the ERs for the interventions identified in phase 1 relative to the ratings for their design components in the 4C/ID model, namely, LT, SI, PTP, and JIT. For an overview of the exact scoring rules, refer to Table 2 in the Methods section. All studies were rated by 2 independent raters. Initial interrater reliability was determined with the intraclass correlation coefficient (ICC). The ICC was 0.969 with a 95% confidence interval 0.950 to 0.981 ($F_{72,2}=33.077$; $P<.01$). A total of 4 cases in which raters 1 and 2 differed were identified and discussed until consensus was reached.

Table 3. Intervention effectiveness and ratings for 4-component instructional design components in phase 1.

Author (year), country	Effectiveness rating	Learning task	Supportive information	Part-task practice	Just-in-time information	Duration (weeks)
Allexandre et al (2016), United States [29]	++[a]	2	2	0	2	8
Boettcher et al (2014), Sweden [30]	++	2	2	2	1	8
Carissoli et al (2015), Italy [31]	+[b]	2	0	0	0	3
Cavanagh et al (2013), United Kingdom [32]	++	2	2	0	2	2
Davis and Zautra (2013), United States [33]	++	2	2	2	0	6
Dimidjian et al (2014), United States [34]	++	1	2	1	0	8
Dowd et al (2015), Ireland [35]	++	2	2	0	2	6
Glück and Maercker (2011), Austria [36]	+	2	0	0	1	2
Gotink et al (2017), The Netherlands [37]	++	2	1	2	2	17
Howells et al (2014), United Kingdom [38]	+	2	1	0	0	1.5
Ly et al (2014), Sweden [39]	+	2	2	0	2	8
Mak et al (2015), China [40]	+	2	2	2	1	8
Michel et al (2014), Germany [41]	++	2	2	0	1	3
Morledge et al (2013), United States [42]	++	2	2	2	2	8
Noguchi et al (2017), Japan [43]	+	1	1	0	1	5
O'Leary and Dockray (2015), Ireland [44]	0[c]	1	0	1	0	3
Querstret et al (2017), United States [45]	++	2	2	2	1	4
Younge et al (2015), The Netherlands [46]	++	2	2	2	2	12

[a]++ indicates that the intervention was rated as more effective.

[b]+ indicates that the intervention was rated as less effective.

[c]0 indicates that the intervention was rated as ineffective.

Table 4. Average ratings for 4-component instructional design components by effectiveness in phase 1.

Intervention code	Average score				
	Learning task	Supportive information	Part-task practice	Just-in-time information	Duration (weeks)
More effective (n=11)	1.91	1.91	1.18	1.35	7.45
Less effective (n=6)	1.83	1.00	0.33	0.83	4.58
Ineffective (n=1)	1.00	1.00	0.00	1.00	3.00

On the basis of the average ratings reported above, we drew the following conclusions that served as the hypothesis for phase 2 of CIS:

- More effective interventions implement formal mindfulness exercises of varying content at least twice per week. Moreover, they continuously provide supportive educational material or reflection exercises, or both. Furthermore, they provide informal PTP opportunities about once per week, provide JIT in the form of reminders at least once per week or when adherence declines, and last for an average of 7 weeks.
- Less effective interventions also implement formal mindfulness exercises of varying content at least twice per week, but SI is only provided once or upon demand. The interventions contain hardly any informal PTP and provide JIT in the form of reminders only about once per week or when adherence declines. The average duration of less effective interventions is 5 weeks.
- Ineffective interventions implement formal mindfulness exercises less than twice per week and provide no SI, no JIT, and hardly any PTP opportunities. The average duration of ineffective interventions is 3 weeks.

Phase 2: Representative Sampling

The search and review strategy for phase 2 is described in detail in the Methods section. The search in phase 2 yielded 14 additional empirical studies that matched our search criteria [47-60]. The 14 studies and their interventions are described in detail further.

General Description and Intervention Effectiveness

The general descriptions of the characteristics of the 14 studies (eg, authors, year of publication, study design, follow-up measures, sample size, age, and indication) included in phase 2 are reported in detail for each study in Multimedia Appendix 4. The quality scores in terms of risk bias are also reported in Multimedia Appendix 4. In summary, 2 studies achieved a quality score of 14, 1 a score of 13, 3 a score of 12, 2 a score of 10, 2 a score of 9, and 1 a score of 7. The vast majority of interventions (n=12) were again aimed at psychological symptoms, such as stress, anxiety, depression, and general well-being. Only 1 intervention was aimed at weight in relation to stress.

A detailed overview of intervention effectiveness (eg, outcome measures, control groups, within- and between-group effects, and computed ERs) is reported in Multimedia Appendix 5. In summary, 6 of the 14 studies were classified as more effective, 6 as less effective, and 2 as ineffective.

Intervention Design and Effectiveness

The intervention design, duration and scheduling, and adherence and acceptance of the interventions are reported in detail in Multimedia Appendix 6. For an overview of intervention design relative to its effectiveness, ERs are reported again.

Intervention Ratings for 4-Component Instructional Design Components

Tables 5 and 6 below report the ERs for the interventions identified in phase 2, relative to the ratings for their design components in the 4C/ID model, namely, LT, SI, PTP, and JIT. For an overview of the exact scoring rules, refer to Table 2. All studies were rated by 2 independent raters. Initial interrater reliability was determined with the ICC. The ICC was 0.918 with a 95% confidence interval from 0.854 to 0.953 ($F_{56,2}=13.167$; $P<.01$). A total of 5 cases in which raters 1 and 2 differed were identified and discussed until consensus was reached.

As the same intervention was implemented across 3 separate studies, an average rating score across those 3 studies was computed for each of the instructional design components. Hence, the number of reported interventions (N=12) does not match the number of reviewed studies (N=14) in Table 6 for phase 2.

Table 5. Intervention effectiveness and ratings for 4-component instructional design components in phase 2.

Author (year), country	Effectiveness rating	Learning task	Supportive information	Part-task practice	Just-in-time information	Duration (weeks)
Antonson et al (2018), Sweden [47]	0	2	1	0	0	8
Bostock et al (2018), United Kingdom[a] [48]	++[b]	2	2	1	2	8
Champion et al (2018), United Kingdom[a] [49]	++	2	2	1	1	4
Joyce et al (2019), Australia [50]	+[c]	2	2	2	2	6
Kvillemo et al (2016), Sweden [51]	+	1	2	2	1	8
Lindsay et al (2018), United States [52]	+	2	1	2	2	2
Lyzwinski et al (2019), Australia [53]	++	2	2	2	2	11
Ma et al (2018), China [54]	+	1	2	1	1	8
Nguyen-Feng et al (2017), United States [55]	+	1	2	0	2	4
Querstret et al (2018), United Kingdom [56]	++	2	2	2	1	4
Shore et al (2018), United Kingdom [57]	++	2	2	2	2	2
van Emmerik et al (2018), The Netherlands [58]	++	2	2	1	2	8
Wahbeh and Oken (2016), United States [59]	0[d]	2	2	2	0	6
Yang et al (2019), United States[a] [60]	++	2	1	1	1	4

[a]Intervention Headspace.

[b]++ indicates that the intervention was rated as more effective.

[c]+ indicates that the intervention was rated as less effective.

[d]0 indicates that the intervention was rated as ineffective.

Table 6. Average ratings for 4-component instructional design components by effectiveness in phase 2.

Intervention code	Average Rating Score for each of the 4 instructional design components				
	Learning task	Supportive information	Part-task practice	Just-in-time information	Duration (weeks)
More effective (n=5)	2.00	1.93	1.60	1.67	5.86
Less effective (n=5)	1.40	1.80	1.40	1.60	5.6
Ineffective (n=2)	2.00	1.50	1.00	0.00	7.00

On the basis of the average ratings reported above, we drew the following conclusions for phase 2 of CIS:

- More effective interventions implement formal mindfulness exercises of varying content at least twice per week. Moreover, they continuously provide supportive educational material or reflection exercises, or both. Furthermore, they provide PTP opportunities implemented at least twice per week, and provide JIT in the form of reminders for each practice, which are sometimes adjustable or contain prompts for self-monitoring. The average duration of more effective interventions in phase 2 is 6 weeks.
- Less effective interventions implement formal mindfulness exercises less than twice per week. Moreover, they continuously provide supportive educational material or reflection exercises, or both. Furthermore, they provide PTP opportunities about once per week, and provide JIT in the form of reminders for each practice, which are sometimes adjustable or contain prompts for

self-monitoring. The average duration of less effective interventions is 6 weeks.

- Ineffective interventions implement formal mindfulness exercises of varying content at least twice per week. Moreover, they continuously provide supportive educational material or reflection exercises, or both. Furthermore, they provide PTP opportunities about once per week, but they provide no JIT. The average duration of ineffective interventions is 7 weeks.

Discussion

Principal Findings

This paper addressed the following 2 research questions:

1. *Which instructional design components can be identified in existing internet-based mindfulness interventions?*
2. *How can these design components be classified relative to the intervention effectiveness?*

With reference to the 4C/ID model [20], the CIS method [26] was applied across 2 phases (diversity sampling and representative sampling) to source for relevant literature. We determined the effectiveness of the identified studies in accordance with the criteria for defining intervention effectiveness (Table 1) [21] and rated the effectiveness of the interventions relative to intervention design components in accordance with the system we developed for this paper (Table 2). Phase 1 yielded 18 studies with 18 different interventions (Multimedia Appendices 1-3); phase 2 yielded an additional 14 studies with 12 different interventions (Multimedia Appendices 4-6). In the 32 studies identified across phases 1 and 2, 5 achieved a risk bias quality score of 14, 10 a score of 13, 6 a score of 12, 3 a score of 10, 4 a score of 9, and 1 a score of 7. The majority of studies (n=24) aimed at psychological symptoms, such as stress, anxiety, depression, or general well-being. A total of 5 studies aimed at physiological symptoms related to fibromyalgia, chronic pain, heart disease, and body weight. A total of 2 interventions aimed at work-life balance and work-related well-being. The 32 studies contained 30 different interventions. Of those interventions, 17 classified as more effective, 12 as less effective, and 3 as ineffective. When comparing the results of phases 1 and 2, the following picture emerges.

More effective interventions consistently implemented formal mindfulness exercises of varying content at least twice per week, continuously provided educational and supportive material and/or reflection exercises, provided informal PTP opportunities at least once per week, and provided JIT in the form of reminders at least once per week or when adherence declined. The average duration of more successful interventions across phases 1 and 2 was 6.5 weeks.

Less effective interventions also consistently implemented formal mindfulness exercises of varying content at least twice per week. These provided SI at least once and contained informal PTP opportunities about once per week or less. JIT in the form of reminders was provided at least once per week or when adherence declined. The average duration of less effective interventions across phases 1 and 2 was 5.5 weeks.

Ineffective interventions implemented formal mindfulness exercises at least once per week but varied strongly in the level of SI they provided. These provided informal PTP opportunities only up to once per week and did not provide any JIT (ie, no reminders). The average duration of ineffective interventions across phases 1 and 2 was 5 weeks.

In summary, the overwhelming majority of internet-based mindfulness interventions are more or less effective, and the effectiveness of the interventions increases with the level of support provided by instructional design components. The difference between effective and ineffective interventions is the presence of JIT in the form of reminders, in addition to the availability of LTs (ie, formal mindfulness exercises), SI (ie, educational material and/or reflection exercises), and PTP (ie, reminders to practice and/or prompts for self-monitoring). We thus conclude that to be effective at all, internet-based mindfulness interventions must contain all 4 4C/ID design components. The difference between more effective and less effective interventions is the presence of continuous support with information in the form of educational materials and/or self-reflection exercises, as compared with SI that is optional or only provided once. The duration of the interventions alone does not seem to be systematically related to intervention effectiveness when taking into account the findings of phases 1 and 2 of our CIS. Phase 1 suggested that ineffective interventions are shorter (average duration 3 weeks) than less and more effective interventions (average durations between 5 and 7 weeks). However, this notion was not supported in phase 2 that revealed no systematic differences in intervention effectiveness based on duration.

Limitations

This CIS is limited, naturally, by its scope and search criteria. For example, only studies in international peer-reviewed journals were considered, and there might be a number of interesting dissertations, conference presentations, and thesis projects, which may contribute to a further understanding of the design components and effectiveness of internet-based mindfulness interventions, that remained unidentified in this review. In addition, there are limitations regarding search terms and sensitivity. As this literature review relied heavily upon relatively broad search criteria such as *meditation*, it attempted to detect studies on internet-based mindfulness interventions with high sensitivity. However, it is likely that a number of studies remained undetected by the applied search strategy. We addressed this issue by extending our original searches with an additional hand search in phase 2, thus identifying an additional 14 studies. In contrast to systematic reviews, the CIS method [26] does not require an exhaustive literature search. Nonetheless, we made efforts to identify all published papers that matched the focus of this investigation.

On the level of the individual studies, limitations relate to small sample sizes in some cases and uneven gender distributions. Regarding statistical analyses, some studies reported only results for per-protocol analyses and not for intention-to-treat analyses that attempt to reduce bias resulting from missing data of dropouts and withdrawals. However, high attrition rates and intention-to-treat analyses may diminish the power to detect

group differences [61]. Risks of bias can also be derived from the requirement for participants to have regular access to an internet-enabled device, which points to the *digital divide*, that is, the gap in internet access between the general population and underserved populations [18] that might be in more need of health-improving interventions [18]. Another potential source for bias concerns the publication bias, that is, the circumstance that predominantly studies with significant effects are published, and studies with nonsignificant results for the same interventions may go unnoticed [21].

In terms of interpreting the results of this CIS, we noticed that it would be insightful to contrast the instructional design components of experimental and active control groups. Although addressing this interesting question exceeds the scope of this investigation, it provides ample incentive and opportunity for future research.

Comparison With Previous Studies

In the past 4 years, 7 major reviews were published on the topic of internet-based mindfulness interventions [9-15]. Those reviews focused on intervention effects and revealed heterogeneous, but predominantly encouraging, results in support of internet-based mindfulness interventions for a variety of mental and physical health conditions [9-15]. Of these 7 reviews, 4 [10,12,14,15] specifically point to design components as the potential sources of variance in intervention effectiveness and call for research to address this issue. This paper serves the purpose.

In line with the existing reviews [9-15], we also found that the majority of the interventions were more or less effective, particularly with regard to mental health. This investigation extends previous studies by providing insight into the instructional design of the implemented interventions relative to the intervention effectiveness. In addition, the rating system for the instructional design components, which we developed based on the 4C/ID model for the purposes of this investigation, is now available to other researchers as a useful tool to classify design components of interventions.

Conclusions

The vast majority of internet-based mindfulness interventions that were identified for this CIS were more or less effective in producing significant changes in the assessed outcome measures. The main difference between effective and ineffective interventions is the presence of JIT in the form of reminders, in addition to the availability of the other 3 design components—LTs, SI, and PTP. The main difference between more effective and less effective interventions is the presence of continuous support with information in the form of educational materials and/or self-reflection exercises, as compared with SI that is optional or only provided once. In summary, we conclude that the effectiveness of the interventions increases with the level of support provided by the instructional design components.

Acknowledgments

The authors acknowledge the studies of the researchers based on which this literature review was conducted.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Characteristics of the included studies in phase 1.
[PDF File (Adobe PDF File), 276 KB - jmir_v21i11e12497_app1.pdf]

Multimedia Appendix 2
Intervention effectiveness of the included studies in phase 1.
[PDF File (Adobe PDF File), 283 KB - jmir_v21i11e12497_app2.pdf]

Multimedia Appendix 3
Intervention design of the included studies in phase 1.
[PDF File (Adobe PDF File), 341 KB - jmir_v21i11e12497_app3.pdf]

Multimedia Appendix 4
Characteristics of the included studies in phase 2.
[PDF File (Adobe PDF File), 267 KB - jmir_v21i11e12497_app4.pdf]

Multimedia Appendix 5
Intervention effectiveness of the included studies in phase 2.
[PDF File (Adobe PDF File), 275 KB - jmir_v21i11e12497_app5.pdf]

Multimedia Appendix 6
Intervention design of the included studies in phase 2.
[PDF File (Adobe PDF File), 312 KB - jmir_v21i11e12497_app6.pdf]

References

1. Bohlmeijer E, Prenger R, Taal E, Cuijpers P. The effects of mindfulness-based stress reduction therapy on mental health of adults with a chronic medical disease: a meta-analysis. J Psychosom Res 2010 Jun;68(6):539-544. [doi: 10.1016/j.jpsychores.2009.10.005] [Medline: 20488270]
2. Grossman P, Niemann L, Schmidt S, Walach H. Mindfulness-based stress reduction and health benefits. A meta-analysis. J Psychosom Res 2004 Jul;57(1):35-43. [doi: 10.1016/S0022-3999(03)00573-7] [Medline: 15256293]
3. Chiesa A. The difficulty of defining mindfulness: current thought and critical issues. Mindfulness 2012 Jun 29;4(3):255-268. [doi: 10.1007/s12671-012-0123-4]
4. Keng S, Smoski MJ, Robins CJ. Effects of mindfulness on psychological health: a review of empirical studies. Clin Psychol Rev 2011 Aug;31(6):1041-1056 [FREE Full text] [doi: 10.1016/j.cpr.2011.04.006] [Medline: 21802619]
5. Kabat-Zinn J. Mindfulness-based interventions in context: past, present, and future. Clin Psychol Sci Prac 2003;10(2):144-156. [doi: 10.1093/clipsy/bpg016]
6. Kabat-Zinn J. Full Catastrophe Living: Using the Wisdom of Your Body and Mind to Face Stress, Pain, and Illness. New York: Delacorte; 1990.
7. Segal ZV, Williams JM, Teasdale JD. Mindfulness-based Cognitive Therapy For Depression: A New Approach To Preventing Relapse. New York: The Guilford Press; 2002.
8. Eichenberg C, Ott R. Klinisch-psychologische Intervention im Internet: Review zu empirischen Befunden störungsspezifischer Angebote. Psychotherapeut 2012;57(1):58-69. [doi: 10.1007/s00278-011-0832-5]
9. Spijkerman MP, Pots WT, Bohlmeijer ET. Effectiveness of online mindfulness-based interventions in improving mental health: a review and meta-analysis of randomised controlled trials. Clin Psychol Rev 2016 Apr;45:102-114 [FREE Full text] [doi: 10.1016/j.cpr.2016.03.009] [Medline: 27111302]
10. Fish J, Brimson J, Lynch S. Mindfulness interventions delivered by technology without facilitator involvement: what research exists and what are the clinical outcomes? Mindfulness (N Y) 2016;7(5):1011-1023 [FREE Full text] [doi: 10.1007/s12671-016-0548-2] [Medline: 27642370]
11. Toivonen KI, Zernicke K, Carlson LE. Web-based mindfulness interventions for people with physical health conditions: systematic review. J Med Internet Res 2017 Aug 31;19(8):e303 [FREE Full text] [doi: 10.2196/jmir.7487] [Medline: 28860106]
12. Heber E, Ebert DD, Lehr D, Cuijpers P, Berking M, Nobis S, et al. The benefit of web- and computer-based interventions for stress: a systematic review and meta-analysis. J Med Internet Res 2017 Feb 17;19(2):e32 [FREE Full text] [doi: 10.2196/jmir.5774] [Medline: 28213341]
13. Sevilla-Llewellyn-Jones J, Santesteban-Echarri O, Pryor I, McGorry P, Alvarez-Jimenez M. Web-based mindfulness interventions for mental health treatment: systematic review and meta-analysis. JMIR Ment Health 2018 Sep 25;5(3):e10278 [FREE Full text] [doi: 10.2196/10278] [Medline: 30274963]
14. Mikolasek M, Berg J, Witt CM, Barth J. Effectiveness of mindfulness- and relaxation-based eHealth interventions for patients with medical conditions: a systematic review and synthesis. Int J Behav Med 2018 Feb;25(1):1-16. [doi: 10.1007/s12529-017-9679-7] [Medline: 28752414]
15. Lyzwinski LN, Caffery L, Bambling M, Edirippulige S. A systematic review of electronic mindfulness-based therapeutic interventions for weight, weight-related behaviors, and psychological stress. Telemed J E Health 2018 Mar;24(3):173-184. [doi: 10.1089/tmj.2017.0117] [Medline: 28885896]
16. Bennett GG, Glasgow RE. The delivery of public health interventions via the internet: actualizing their potential. Annu Rev Public Health 2009;30:273-292. [doi: 10.1146/annurev.publhealth.031308.100235] [Medline: 19296777]
17. Cuijpers P, van Straten A, Andersson G. Internet-administered cognitive behavior therapy for health problems: a systematic review. J Behav Med 2008 Apr;31(2):169-177 [FREE Full text] [doi: 10.1007/s10865-007-9144-1] [Medline: 18165893]
18. Ybarra ML, Eaton WW. Internet-based mental health interventions. Ment Health Serv Res 2005 Jun;7(2):75-87. [Medline: 15974154]
19. Hilgart MM, Ritterband LM, Thorndike FP, Kinzie MB. Using instructional design process to improve design and development of internet interventions. J Med Internet Res 2012 Jun 28;14(3):e89 [FREE Full text] [doi: 10.2196/jmir.1890] [Medline: 22743534]
20. van Merriënboer JJ, Clark RE, de Croock MB. Blueprints for complex learning: The 4C/ID-model. Educ Technol Res Dev 2002;50(2):39-61. [doi: 10.1007/BF02504993]
21. Morrison LG, Yardley L, Powell J, Michie S. What design features are used in effective e-Health interventions? A review using techniques from critical interpretive synthesis. Telemed J E Health 2012 Mar;18(2):137-144. [doi: 10.1089/tmj.2011.0062] [Medline: 22381060]

22. Lustria ML, Cortese J, Noar SM, Glueckauf RL. Computer-tailored health interventions delivered over the web: review and analysis of key components. Patient Educ Couns 2009 Feb;74(2):156-173. [doi: 10.1016/j.pec.2008.08.023] [Medline: 18947966]
23. Kelders SM, Kok RN, Ossebaard HC, van Gemert-Pijnen JE. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012 Nov 14;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]
24. Eysenbach G, Powell J, Englesakis M, Rizo C, Stern A. Health related virtual communities and electronic support groups: systematic review of the effects of online peer to peer interactions. Br Med J 2004 May 15;328(7449):1166 [FREE Full text] [doi: 10.1136/bmj.328.7449.1166] [Medline: 15142921]
25. Cavanagh K, Strauss C, Forder L, Jones F. Can mindfulness and acceptance be learnt by self-help?: a systematic review and meta-analysis of mindfulness and acceptance-based self-help interventions. Clin Psychol Rev 2014 Mar;34(2):118-129. [doi: 10.1016/j.cpr.2014.01.001] [Medline: 24487343]
26. Dixon-Woods M, Cavers D, Agarwal S, Annandale E, Arthur A, Harvey J, et al. Conducting a critical interpretive synthesis of the literature on access to healthcare by vulnerable groups. BMC Med Res Methodol 2006 Jul 26;6:35 [FREE Full text] [doi: 10.1186/1471-2288-6-35] [Medline: 16872487]
27. van Tulder M, Furlan A, Bombardier C, Bouter L, Editorial Board of the Cochrane Collaboration Back Review Group. Updated method guidelines for systematic reviews in the cochrane collaboration back review group. Spine (Phila Pa 1976) 2003 Jun 15;28(12):1290-1299. [doi: 10.1097/01.BRS.0000065484.95996.AF] [Medline: 12811274]
28. Mars TS, Abbey H. Mindfulness meditation practise as a healthcare intervention: a systematic review. Int J Osteopath Med 2010;13(2):56-66. [doi: 10.1016/j.ijosm.2009.07.005]
29. Allexandre D, Bernstein AM, Walker E, Hunter J, Roizen MF, Morledge TJ. A web-based mindfulness stress management program in a corporate call center: a randomized clinical trial to evaluate the added benefit of onsite group support. J Occup Environ Med 2016 Mar;58(3):254-264 [FREE Full text] [doi: 10.1097/JOM.0000000000000680] [Medline: 26949875]
30. Boettcher J, Aström V, Påhlsson D, Schenström O, Andersson G, Carlbring P. Internet-based mindfulness treatment for anxiety disorders: a randomized controlled trial. Behav Ther 2014 Mar;45(2):241-253 [FREE Full text] [doi: 10.1016/j.beth.2013.11.003] [Medline: 24491199]
31. Carissoli C, Villani D, Riva G. Does a meditation protocol supported by a mobile application help people reduce stress? Suggestions from a controlled pragmatic trial. Cyberpsychol Behav Soc Netw 2015 Jan;18(1):46-53. [doi: 10.1089/cyber.2014.0062] [Medline: 25584730]
32. Cavanagh K, Strauss C, Cicconi F, Griffiths N, Wyper A, Jones F. A randomised controlled trial of a brief online mindfulness-based intervention. Behav Res Ther 2013 Sep;51(9):573-578. [doi: 10.1016/j.brat.2013.06.003] [Medline: 23872699]
33. Davis MC, Zautra AJ. An online mindfulness intervention targeting socioemotional regulation in fibromyalgia: results of a randomized controlled trial. Ann Behav Med 2013 Dec;46(3):273-284. [doi: 10.1007/s12160-013-9513-7] [Medline: 23670111]
34. Dimidjian S, Beck A, Felder JN, Boggs JM, Gallop R, Segal ZV. Web-based mindfulness-based cognitive therapy for reducing residual depressive symptoms: an open trial and quasi-experimental comparison to propensity score matched controls. Behav Res Ther 2014 Dec;63:83-89 [FREE Full text] [doi: 10.1016/j.brat.2014.09.004] [Medline: 25461782]
35. Dowd H, Hogan MJ, McGuire BE, Davis MC, Sarma KM, Fish RA, et al. Comparison of an online mindfulness-based cognitive therapy intervention with online pain management psychoeducation: a randomized controlled study. Clin J Pain 2015 Jun;31(6):517-527. [doi: 10.1097/AJP.0000000000000201] [Medline: 25565584]
36. Glück TM, Maercker A. A randomized controlled pilot study of a brief web-based mindfulness training. BMC Psychiatry 2011 Nov 8;11:175 [FREE Full text] [doi: 10.1186/1471-244X-11-175] [Medline: 22067058]
37. Gotink RA, Younge JO, Wery MF, Utens EM, Michels M, Rizopoulos D, et al. Online mindfulness as a promising method to improve exercise capacity in heart disease: 12-month follow-up of a randomized controlled trial. PLoS One 2017;12(5):e0175923 [FREE Full text] [doi: 10.1371/journal.pone.0175923] [Medline: 28486559]
38. Howells A, Ivtzan I, Eiroa-Orosa FJ. Putting the 'app' in happiness: a randomised controlled trial of a smartphone-based mindfulness intervention to enhance wellbeing. J Happiness Stud 2014;17(1):163-185. [doi: 10.1007/s10902-014-9589-1]
39. Ly KH, Trüschel A, Jarl L, Magnusson S, Windahl T, Johansson R, et al. Behavioural activation versus mindfulness-based guided self-help treatment administered through a smartphone application: a randomised controlled trial. BMJ Open 2014 Jan 9;4(1):e003440 [FREE Full text] [doi: 10.1136/bmjopen-2013-003440] [Medline: 24413342]
40. Mak WW, Chan AT, Cheung EY, Lin CL, Ngai KC. Enhancing web-based mindfulness training for mental health promotion with the health action process approach: randomized controlled trial. J Med Internet Res 2015 Jan 19;17(1):e8 [FREE Full text] [doi: 10.2196/jmir.3746] [Medline: 25599904]
41. Michel A, Bosch C, Rexroth M. Mindfulness as a cognitive-emotional segmentation strategy: an intervention promoting work-life balance. J Occup Organ Psychol 2014;87(4):733-754. [doi: 10.1111/joop.12072]
42. Morledge TJ, Allexandre D, Fox E, Fu AZ, Higashi MK, Kruzikas DT, et al. Feasibility of an online mindfulness program for stress management--a randomized, controlled trial. Ann Behav Med 2013 Oct;46(2):137-148 [FREE Full text] [doi: 10.1007/s12160-013-9490-x] [Medline: 23632913]

43. Noguchi R, Sekizawa Y, So M, Yamaguchi S, Shimizu E. Effects of five-minute internet-based cognitive behavioral therapy and simplified emotion-focused mindfulness on depressive symptoms: a randomized controlled trial. BMC Psychiatry 2017 Mar 4;17(1):85 [FREE Full text] [doi: 10.1186/s12888-017-1248-8] [Medline: 28259151]
44. O'Leary K, Dockray S. The effects of two novel gratitude and mindfulness interventions on well-being. J Altern Complement Med 2015 Apr;21(4):243-245. [doi: 10.1089/acm.2014.0119] [Medline: 25826108]
45. Querstret D, Cropley M, Fife-Schaw C. Internet-based instructor-led mindfulness for work-related rumination, fatigue, and sleep: assessing facets of mindfulness as mechanisms of change. A randomized waitlist control trial. J Occup Health Psychol 2017 Apr;22(2):153-169. [doi: 10.1037/ocp0000028] [Medline: 27054503]
46. Younge JO, Wery MF, Gotink RA, Utens EM, Michels M, Rizopoulos D, et al. Web-based mindfulness intervention in heart disease: a randomized controlled trial. PLoS One 2015;10(12):e0143843 [FREE Full text] [doi: 10.1371/journal.pone.0143843] [Medline: 26641099]
47. Antonson C, Thorsén F, Sundquist J, Sundquist K. Upper secondary school students' compliance with two internet-based self-help programmes: a randomised controlled trial. Eur Child Adolesc Psychiatry 2018 Feb;27(2):191-200 [FREE Full text] [doi: 10.1007/s00787-017-1035-6] [Medline: 28776094]
48. Bostock S, Crosswell AD, Prather AA, Steptoe A. Mindfulness on-the-go: effects of a mindfulness meditation app on work stress and well-being. J Occup Health Psychol 2019 Feb;24(1):127-138. [doi: 10.1037/ocp0000118] [Medline: 29723001]
49. Champion L, Economides M, Chandler C. The efficacy of a brief app-based mindfulness intervention on psychosocial outcomes in healthy adults: a pilot randomised controlled trial. PLoS One 2018;13(12):e0209482 [FREE Full text] [doi: 10.1371/journal.pone.0209482] [Medline: 30596696]
50. Joyce S, Shand F, Lal TJ, Mott B, Bryant RA, Harvey SB. Resilience@Work mindfulness program: results from a cluster randomized controlled trial with first responders. J Med Internet Res 2019 Feb 19;21(2):e12894 [FREE Full text] [doi: 10.2196/12894] [Medline: 30777846]
51. Kvillemo P, Brandberg Y, Bränström R. Feasibility and outcomes of an internet-based mindfulness training program: a pilot randomized controlled trial. JMIR Ment Health 2016 Jul 22;3(3):e33 [FREE Full text] [doi: 10.2196/mental.5457] [Medline: 27450466]
52. Lindsay EK, Young S, Smyth JM, Brown KW, Creswell JD. Acceptance lowers stress reactivity: dismantling mindfulness training in a randomized controlled trial. Psychoneuroendocrinology 2018 Jan;87:63-73. [doi: 10.1016/j.psyneuen.2017.09.015] [Medline: 29040891]
53. Lyzwinski LN, Caffery L, Bambling M, Edirippulige S. The mindfulness app trial for weight, weight-related behaviors, and stress in university students: randomized controlled trial. JMIR Mhealth Uhealth 2019 Apr 10;7(4):e12210 [FREE Full text] [doi: 10.2196/12210] [Medline: 30969174]
54. Ma Y, She Z, Siu AF, Zeng X, Liu X. Effectiveness of online mindfulness-based interventions on psychological distress and the mediating role of emotion regulation. Front Psychol 2018;9:2090 [FREE Full text] [doi: 10.3389/fpsyg.2018.02090] [Medline: 30429816]
55. Nguyen-Feng VN, Greer CS, Frazier P. Using online interventions to deliver college student mental health resources: evidence from randomized clinical trials. Psychol Serv 2017 Nov;14(4):481-489. [doi: 10.1037/ser0000154] [Medline: 29120206]
56. Querstret D, Cropley M, Fife-Schaw C. The effects of an online mindfulness intervention on perceived stress, depression and anxiety in a non-clinical sample: a randomised waitlist control trial. Mindfulness (N Y) 2018;9(6):1825-1836 [FREE Full text] [doi: 10.1007/s12671-018-0925-0] [Medline: 30524514]
57. Shore R, Strauss C, Cavanagh K, Hayward M, Ellett L. A randomised controlled trial of a brief online mindfulness-based intervention on paranoia in a non-clinical sample. Mindfulness (N Y) 2018;9(1):294-302 [FREE Full text] [doi: 10.1007/s12671-017-0774-2] [Medline: 29387267]
58. van Emmerik AA, Berings F, Lancee J. Efficacy of a mindfulness-based mobile application: a randomized waiting-list controlled trial. Mindfulness (N Y) 2018;9(1):187-198 [FREE Full text] [doi: 10.1007/s12671-017-0761-7] [Medline: 29387266]
59. Wahbeh H, Oken BS. Internet mindfulness meditation intervention for the general public: pilot randomized controlled trial. JMIR Ment Health 2016 Aug 8;3(3):e37 [FREE Full text] [doi: 10.2196/mental.5900] [Medline: 27502759]
60. Yang E, Schamber E, Meyer RM, Gold JI. Happier healers: randomized controlled trial of mobile mindfulness for stress management. J Altern Complement Med 2018 May;24(5):505-513. [doi: 10.1089/acm.2015.0301] [Medline: 29420050]
61. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]

Abbreviations

4C/ID: 4-component instructional design
CIS: critical interpretive synthesis
ER: effectiveness rating
ICC: intraclass correlation coefficient

JIT: just-in-time information
LT: learning task
PTP: part-task practice
SI: supportive information

Edited by G Eysenbach; submitted 12.10.18; peer-reviewed by K Cavanagh, K Douw, N Farič; comments to author 21.03.19; revised version received 17.08.19; accepted 26.09.19; published 27.11.19.

Please cite as:
Lippmann M, Laudel H, Heinzle M, Narciss S
Relating Instructional Design Components to the Effectiveness of Internet-Based Mindfulness Interventions: A Critical Interpretive Synthesis
J Med Internet Res 2019;21(11):e12497
URL: http://www.jmir.org/2019/11/e12497/
doi:10.2196/12497
PMID:31774413

Review

Effects of Mobile Health App Interventions on Sedentary Time, Physical Activity, and Fitness in Older Adults: Systematic Review and Meta-Analysis

Dharani Yerrakalva[1,2], BSc, MBBS, MPhil; Dhrupadh Yerrakalva[3], BSc, MBChB; Samantha Hajna[2], BSc, MSc, PhD; Simon Griffin[1], BSc, MBBS, MSc, MD

[1]Primary Care Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, United Kingdom

[2]Medical Research Council Epidemiology Unit, University of Cambridge, Cambridge, United Kingdom

[3]Barking, Havering, and Redbridge University Hospitals Trust, London, United Kingdom

Corresponding Author:
Dharani Yerrakalva, BSc, MBBS, MPhil
Primary Care Unit
Department of Public Health and Primary Care
University of Cambridge
Institute of Public Health
Cambridge,
United Kingdom
Phone: 44 1223 330300
Email: dharaniyerrakalva@googlemail.com

Abstract

Background: High sedentary time, low physical activity (PA), and low physical fitness place older adults at increased risk of chronic diseases, functional decline, and premature mortality. Mobile health (mHealth) apps, apps that run on mobile platforms, may help promote active living.

Objective: We aimed to quantify the effect of mHealth app interventions on sedentary time, PA, and fitness in older adults.

Methods: We systematically searched five electronic databases for trials investigating the effects of mHealth app interventions on sedentary time, PA, and fitness among community-dwelling older adults aged 55 years and older. We calculated pooled standardized mean differences (SMDs) in these outcomes between the intervention and control groups after the intervention period. We performed a Cochrane risk of bias assessment and Grading of Recommendations, Assessment, Development, and Evaluation certainty assessment.

Results: Overall, six trials (486 participants, 66.7% [324/486] women; age mean 68 [SD 6] years) were included (five of these trials were included in the meta-analysis). mHealth app interventions may be associated with decreases in sedentary time (SMD=−0.49; 95% CI −1.02 to 0.03), increases in PA (506 steps/day; 95% CI −80 to 1092), and increases in fitness (SMD=0.31; 95% CI −0.09 to 0.70) in trials of 3 months or shorter and with increases in PA (753 steps/day; 95% CI −147 to 1652) in trials of 6 months or longer. Risk of bias was low for all but one study. The quality of evidence was moderate for PA and sedentary time and low for fitness.

Conclusions: mHealth app interventions have the potential to promote changes in sedentary time and PA over the short term, but the results did not achieve statistical significance, possibly because studies were underpowered by small participant numbers. We highlight a need for larger trials with longer follow-up to clarify if apps deliver sustained clinically important effects.

(J Med Internet Res 2019;21(11):e14343) doi:10.2196/14343

KEYWORDS

sedentary behavior; physical activity; physical fitness; aged; mHealth; mobile apps

Introduction

Background

Older adults spend an average 9.4 hours of their day being sedentary [1] and are not meeting current physical activity (PA) recommendations [2,3]. High sedentary time, low PA, and low fitness levels place older adults at increased risk of chronic diseases [4-10], declines in functional and cognitive health [11-13], sarcopenia [14], and premature mortality [15]. Interventions to reduce sedentary time, increase PA, and improve fitness could potentially enhance the health and well-being of older adults. However, sustained positive changes in PA [16-18] and sedentary time [19,20] beyond 12 months have not been consistently achieved through traditional interventions.

Mobile health (mHealth) apps, software apps that run on a mobile platform such as a mobile phone or tablet [21], may provide an alternative approach as they circumvent many of the limitations of traditional professional-led interventions. Common traditional interventions include advice from health care professionals and educational materials. The limitations of these methods are restricted access to professionals and limited reach. Furthermore, there is considerable cost associated with training and employing the professionals required to deliver these interventions. The duration of monitoring and feedback may also be limited by cost and time restraint, and therefore, there is no provision for frequent or any follow-up. Other traditional methods such as educational materials lack tailoring.

Apps have the potential to be more cost-effective than traditional interventions [22]; although this field is developing, cost-effectiveness analyses are not widely available. They can overcome barriers to accessing health care as they can be used independent of a health care provider, and they appeal to healthy individuals and those with medical conditions and therefore have the potential to reach a larger percentage of the population. In addition, app technologies allow GPS monitoring, tailored feedback, and reminders throughout the day. In 2017, there were 325,000 health apps available in major app stores [23]. Older adults are traditionally not seen as app users, but mobile phone use among older adults is significant and increasing [24-26], as is the use of health apps.

mHealth is defined as health interventions involving mobile devices [1], and apps are one such intervention. mHealth interventions themselves are a subset of electronic health (eHealth; an umbrella term that includes any information and communication technologies utilized in the health care field). mHealth apps confer unique advantages over other eHealth (static computer, internet, landline use, text messaging, and mobile telephone calls) as they can allow continuous monitoring that provides the basis for individualized feedback and goal setting.

Health-related apps are currently recommended in a range of health contexts. This includes the management of a variety of health conditions such as depression, dementia, diabetes, and chronic obstructive pulmonary disease [27]; medication adherence and rehabilitation; use as symptoms checkers; and managing clinical records. Apps can provide quick advice, support from peers, and the ability to self-monitor. There is a growing interest in using apps to modify behaviors such as PA or sedentariness to improve or maintain health. These apps commonly feature self-monitoring of behavior, either through self-inputting data or by linking to a monitoring device (eg, pedometer or smart watches). With these data, the apps can provide feedback, prompts, goal setting, rewards, and social connectivity.

Two recent systematic reviews investigated the effectiveness of eHealth interventions on PA [28,29], although no meta-analyses were done. Both reviews reported that eHealth interventions promote PA in the short-term among adults older than 55 years and that evidence regarding the long-term effects is still lacking.

Objectives

There is no existing review examining the specific role of mHealth app interventions on PA in older adults. Furthermore, there is no existing systematic review or meta-analysis examining the effect of mHealth app interventions on sedentary time or fitness in older adults.

Such a review is needed to inform the development of scalable and effective activity interventions among older adults. To fill this gap in knowledge, we aimed to synthesize the existing evidence on the effectiveness of mHealth app interventions on sedentary time, PA, and fitness in older adults and to identify common behavioral change techniques (BCTs) utilized in effective interventions.

Methods

Protocol and Eligibility Criteria

We published a Prospective Register of Systematic Reviews (PROSPERO) protocol before undertaking this review (CRD42018106195). We included trials (randomized or nonrandomized) that compared the effectiveness of an mHealth app intervention with either a modified dose of intervention (modified volume of intervention or modified version of same app), different app, nonapp intervention, or no intervention. To be eligible, the trials had to include community-dwelling adults aged 55 years and older. The outcomes assessed in this study were PA (moderate- to vigorous-intensity physical activity [MVPA] measured by accelerometer and steps/day measured by pedometer), physical fitness (maximal oxygen uptake [VO_2 max], gait speed, and 6-min walk), or sedentary time (sedentary time measured by accelerometer). The outcome measures had to be objectively assessed. We only included trials in the meta-analysis that reported outcome values pre- and postintervention. We limited the searches to human studies published after 2008 as the emergence of mHealth apps occurred after this time. We excluded trials that included nonapp interventions only (ie, other types of eHealth or mHealth interventions only and no apps).

Information Sources and Search

We systematically searched 5 medical electronic databases for trials investigating effects of mHealth app interventions on

sedentary time, PA, and fitness among community-dwelling older adults aged 55 years and older. These included MEDLINE via OVID, PsycINFO via Ovid, Web of Science, Cochrane Central Register of Controlled Trials, and Physical Education Index. We identified key search terms and further expanded this list by running medical subject headings searches. Search terms were classified under 3 main headings: age (aged, elderly, old, and senior), intervention (application, M-health, "mobile health", and e-health), and outcome (sedentary, sitting, physical activity, steps, VO_2 max, exercise, fitness, and functional aerobic capacity). We then conducted Boolean searches to systematically tie the clustered terms (and their variations through truncation) to identify potential articles (example strategy for MEDLINE in Multimedia Appendix 1). We searched for additional papers in the reference lists of review articles, protocols, key commentary articles, and the final included papers.

Study Selection

Two authors (DAY and DRY) independently carried out all steps of the study selection and data collection detailed below. This included screening the titles and abstracts of the search results to identify articles that met the inclusion criteria and retrieving the full text of identified articles. The full-text articles were then screened. We used the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines (flow diagram and checklist) to ensure that we adhered to a high standard of reporting [30].

Data Collection

We extracted data using the Cochrane Public Health Group Data Extraction and Assessment Template [31]. This included study information, participant characteristics, intervention information (duration, intensity, setting, and BCTs employed), outcomes, and risk of bias assessment information.

The BCT taxonomy [32] is used to classify BCTs employed in interventions. It was created as an *agreed language* of 93 distinct BCTs that could be used to describe the *active ingredients* in interventions. Two authors (DAY and DRY) independently underwent the Web-based training to be able to identify BCTs within interventions and then utilized this training to identify BCTs in the included studies.

Data Analysis

For each included study, we calculated the mean difference between the postintervention outcome values of the intervention arm and the control arm. To account for the heterogeneity in the units of the outcome measures, we calculated the standardized mean difference (SMD) between the postintervention values of the intervention arm and the control. We utilized a random-effects model to estimate the pooled effect of interventions on PA, fitness, and sedentary time. Where a study reported more than one outcome, we included the measure that was most homogenous to the other included studies. We used STATA 15.0 for all analyses (StataCorp LP).

To assess heterogeneity, we calculated I^2 values and visually inspected forest plots. We did not examine funnel plots to examine for publication bias given that SMDs are naturally correlated with their standard errors and can produce spurious asymmetry. Furthermore, when there are less than 10 studies included in the meta-analysis, the power of the tests is too low to distinguish chance from real asymmetry.

Risk of Bias Assessment

We assessed the internal validity of each study using the Cochrane Collaboration's tool for assessing risk of bias [33]. Studies were categorized as having high, low, or undetermined risk of bias.

Certainty Assessment

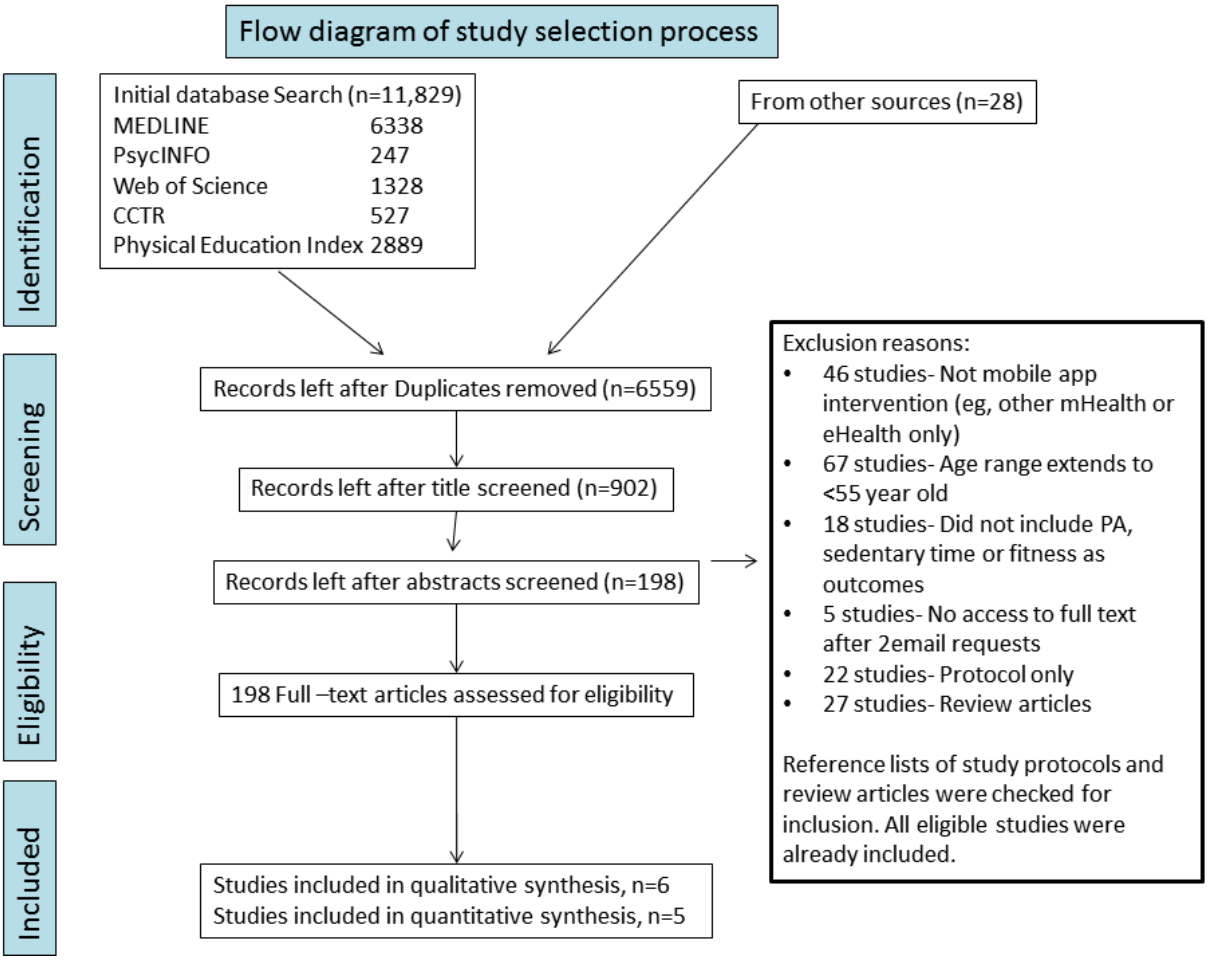
The results from the meta-analysis and risk of bias assessment were used to complete a Grading of Recommendations, Assessment, Development, and Evaluation (GRADE) certainty assessment [34].

Results

Study Selection

There were no discrepancies between the articles retrieved or extracted data by the 2 authors. A total of 11,829 study titles were identified, with 6559 left once duplicates were removed (Figure 1). We assessed 198 full-text articles for eligibility against the inclusion criteria, with reasons listed in Figure 1. Overall, 6 studies were eligible for inclusion in the systematic review, 5 of which were eligible for inclusion in the meta-analysis.

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow diagram. eHealth: electronic health; mHealth: mobile health; PA: physical activity; CCTR: Cochrane Controlled Trial Register.



Study Characteristics

The included articles were published between January 2013 and March 2017. Across the 6 studies, the number of participants in each study ranged from 19 to 263 (mean 81, total N=486; Table 1). In addition, 5 studies used a parallel group randomized controlled trial (RCT) design [35-38], and 1 study was nonrandomized [39]. Studies were conducted in the United States [36,37], Switzerland [39], and Canada [35,38,40]. Participants were aged on average 68 years (SD 6), and 66.7% (324/486) of them were female. The duration of the interventions ranged from 2 to 6 months and the duration of follow-up ranged from 2 to 12 months (Table 1). None of the studies specified whether the apps, or other elements of interventions such as fitness trackers, were taken away from participants at the end of the monitored intervention or left with them to be used until follow-up.

Table 1. Characteristics of included studies.

First author, year	Total number of participants	Number in intervention group	Number in comparator group	Age (years), mean (SD)	Sex female, n (%)	Country	Study design	Duration of intervention (months)	Follow-up measurement (months)
Ashe, 2015 [35]	19	12	7	64 (4.6)	19 (100)	Canada	2-arm RCT[a]	6	3 and 6
Bickmore, 2013 [36]	263	132	131	71.3 (5.4)	161 (61.2)	United States	2-arm RCT	2	2 and 12
Silveira, 2013 [39]	44	Ind[b] 14; Soc[c] 13	17	Ind 74 (5); Soc 75 (6); Cont[d] 76 (15)	Ind 10 (71); Soc 8 (62); Cont 10 (59)	Switzerland	2-arm non-RCT	3	3
Knight, 2014 [40]	60	SB[e] 14; EX[f] 15; CC[g] 16	15	63 (4)	39 (65)	Canada	3-arm random, no control	3	3
Lyons, 2017 [37]	40	20	20	61.5 (5.6)	34 (85)	United States	2-arm RCT	3	3
Knight, 2014 [38]	60	SB 14; EX 15; CC 16	15	63 (4)	39 (65)	Canada	4-arm RCT	3	3

[a]RCT: randomized controlled trial.

[b]Ind: individual intervention group.

[c]Soc: social intervention group.

[d]Cont: control group.

[e]SB: sedentary behavior intervention.

[f]EX: physical activity intervention.

[g]CC: combined intervention.

All app interventions but one involved syncing the app to wearable technology (Table 2). Overall, 3 studies had users syncing or inputting data from pedometers [36,38,40] to apps, and 2 studies [35,37] had users syncing a wearable smart device (Fitbit [Fitbit Inc] and UP24 [Jawbone]) with apps. In 4 studies [35-37,39], app functionality allowed goal setting and tailored feedback or prompts related to progress to goals. In addition, 2 studies included apps that only allowed monitoring of step count, blood pressure, and blood glucose with no goal setting or prompts [38,40].

mHealth app interventions were delivered through smartphones in 3 studies [35,38,40] and tablet computers in 3 studies [36,37,39]. The app was the primary focus of the intervention in 3 studies [36,37,39], whereas the app was used in combination with educational classes and phone calls with health care professionals in 3 studies [35,38,40]. In addition, 2 studies had a *no-content* comparator group [37,40] (Table 2), 2 studies had *nontechnology* comparator groups [35,39], and 1 study had a *technology nonapp* comparator group [36]. In 1 study [38], all groups included mHealth app interventions, and therefore, this study was not included in the meta-analysis. Of the 6 studies, 4 reported the effectiveness of their intervention on PA [35-38], 2 on sedentary time [35,37], and 3 on physical fitness [37,39,40].

Table 2. Intervention and comparator group characteristics.

First author, year	Outcome (measurement)	Intervention description	App characteristics	Comparator description	Intervention frequency
Ashe, 2015 [35]	PA[a] (minutes/day, Actigraph), sedentary time (% sedentary time/day, Actigraph)	Group education, individualized PA prescription, Fitbit with Fitbit app use	App syncs with wearable device. Can view daily steps, calories burned, active minutes, and sleep and track food and weight. Allows monitoring trends, goal setting and tailored daily prompts. Social networking forums	Educational sessions without PA component	4 weekly sessions, then 5 monthly sessions
Bickmore, 2013 [36]	PA (steps/day, pedometer)	Tablet with ECA[b] app and pedometer	App syncs with pedometer. Can view step counts. Has *virtual coach.* Allows monitoring trends and goal setting, identifies barriers, and negotiates new goals. Gives exercise tip of the day.	Pedometer and self-monitoring	Instructed to have 1 conversation with ECA per day
Silveira, 2013 [39]	Fitness (m/s, fastest gait speed)	Introductory class with iPad and active lifestyle app (either *individual* or *social* version)	App has strength-balance training plans, with videos to support exercises. Sends praise/reward messages if there is progress in goals (eg, has a flower that grows if session is completed). Allows monitoring trends. Bulletin board to discuss progress with experts and friends in social version.	Physical exercise manual and paper log	Patient guided
Knight, 2014 [40]	PA (steps/day, pedometer)	Introduction and prescription of PA (exercise, sedentary, or both) then use of smartphone + app, pedometer, glucometer, blood pressure monitor	App syncs with blood pressure and blood glucose monitors. User manually inputs steps/day from pedometer. Allows monitoring trends	No comparator, all groups had apps	Patient guided
Lyons, 2017 [37]	PA (minutes/day, activPAL); steps/day, (pedometer), sedentary time (sitting time/day, activPAL), fitness (meters, 6-min timed walk)	iPad and app/UP24 Jawbone wearable device. Initial visit then telephone counseling	Syncs with wearable device. Allows monitoring trends of step count, heart rate, sleep, food, and weight. Has a *smart coach* that offers tailored advice based on progress, goal setting, and prompts.	Control, no intervention *wait list*	Weekly telephone counseling
Knight, 2014 [38]	Fitness (maximal oxygen uptake)	Introduction and prescription for a specific intensity of PA (exercise, sedentary, or both) then use of smartphone + app, pedometer, glucometer, blood pressure monitor)	App syncs with blood pressure and blood glucose monitors. User manually inputs steps/day from pedometer. Allows monitoring trends	Control, no intervention	Patient guided

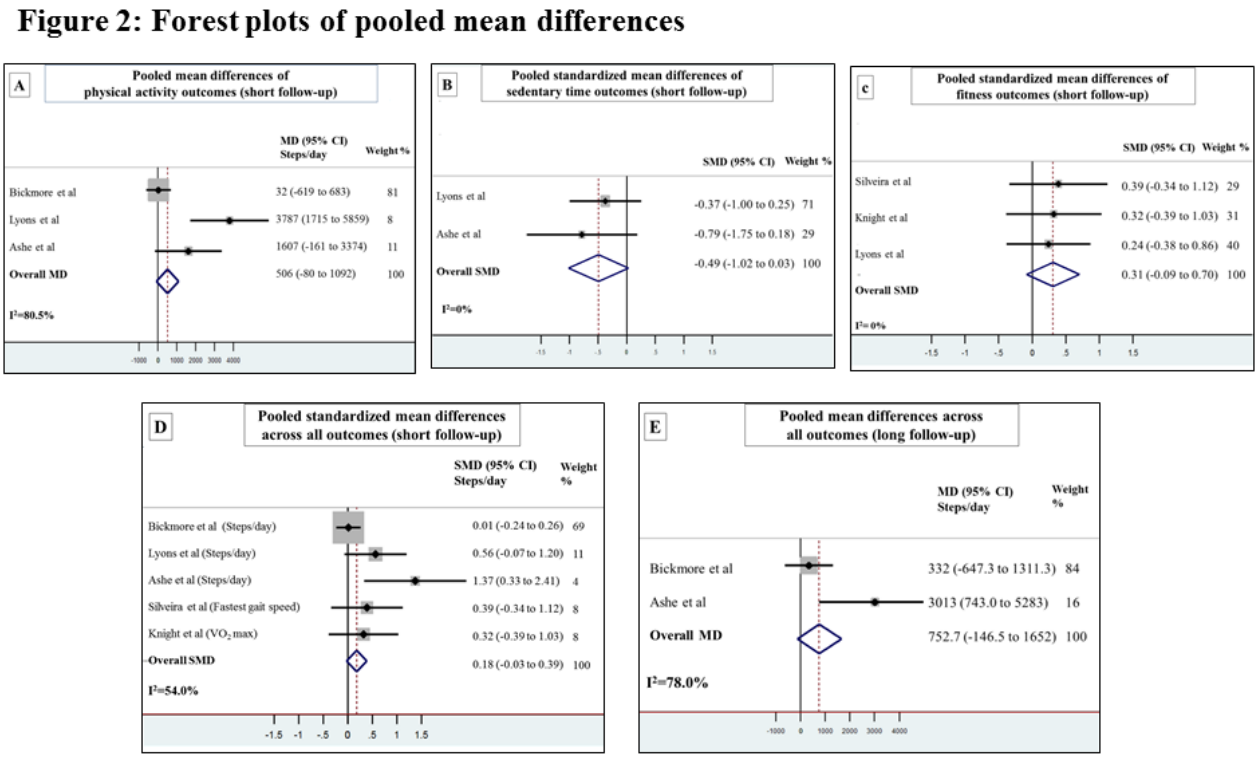
[a]PA: physical activity.

[b]ECA: embodied conversational agent.

Effectiveness of Interventions

The pooled SMD across sedentary time, PA, and fitness outcomes was 0.18 (95% CI −0.03 to 0.39; Figure 2). There was evidence of statistical heterogeneity (I^2=54%).

Figure 2. Forest plots of pooled mean differences . MD: mean difference; SMD: standardized mean difference; VO2 max: maximal oxygen uptake.



Panel A: Outcome measures for all studies were steps/day.
Panel B: Outcome measure for Lyons et al was sitting time/day, and for Ashe et al was % sedentary time/day.
Panel C: Outcome measures for Silveira et al was fastest gait speed, for Knight et al was VO_2 max, and for Lyons et al was 6-min timed walk.
Panel D: Outcome measures for Bickmore et al, Lyons et al and Ashe et al were steps/day, for Silveira et al was fastest gait speed and for Knight et al was VO2 max.
Panel E: Outcome measures for all studies were steps/day.

Sedentary Time

Shorter-Term Effects (≤3 Months)

The pooled SMD for sedentary time across the 2 studies was −0.49 (95% CI −1.02 to 0.03) with no statistical heterogeneity (I^2=0%). The effect of mHealth app interventions was inconclusive for the studies by Lyons et al [37] (−60.5 min/day in sitting time; 95% CI −161 to 40) and Ashe et al [35] (−5.1% sitting time per day; 95% CI −10.8 to 0.31).

Longer-Term Effects (≥6 Months)

Ashe et al [35] reported an inconclusive effect on percentage sedentary time per day over 6 months of follow-up (−1.06% sitting time/day; 95% CI −6.35 to 4.23).

Physical Activity

Shorter-Term Effects (≤3 Months)

mHealth app interventions led to an average increase of 506 steps/day (pooled mean difference 95% CI −80 to 1092) across the 3 studies reporting this outcome [35-37]. Although all of the individual effect estimates were in the same direction, the effect estimates ranged greatly (from 32 to 3787 steps/day), and the pooled effect did not reach statistical significance. Statistical heterogeneity was high (I^2=80.5%). Ashe et al [35] reported 2 measures of PA; but we only included steps/day in the pooled mean difference estimate to maximize homogeneity. Ashe et al reported a 22.1 min/day increase in MVPA (95% CI 6.64 to 37.5) [35].

The study by Knight et al [38] was not included in the pooled analysis as there was no control group. In their study, the authors examined the effect of 3 different interventions (1 intervention targeting sedentary time, 1 targeting PA, and 1 targeting both) on steps/day. The effects of all 3 interventions were inconclusive. Mean changes from baseline to 3 months were 460 steps/day (95% CI −278 to 1199) for the sedentary time intervention, −76 steps/day (95% CI −791 to 640) for the PA intervention, and −454 steps/day (95% CI −1134 to 225) for the combined intervention.

Longer-Term Effects (6-12 Months)

mHealth app interventions led to an average increase of 753 steps/day (pooled mean difference 95% CI −147 to 1652) across the 2 studies reporting this outcome [35,36] (Figure 2). The level of heterogeneity was high (I^2=78%). Ashe et al [35] reported a mean increase of 3013 steps/day (95% CI 743 to 5283) and 19.6 min/day increase in MVPA at 6 months (95% CI 2.2 to 36.9). Bickmore et al [36] reported a mean increase of 332 steps/day (95% CI −647 to 1311) at 12 months.

Physical Fitness

Shorter-Term Effects (≤3 Months)

The pooled SMD for fitness across the 3 studies was 0.31 (95% CI −0.09 to 0.70) with no evidence of statistical heterogeneity (I^2=0%). The individual studies reported that their app interventions had mixed effects on fitness. Silveira et al [39] reported that their intervention led to a 0.47 m/s increase in fastest gait speed (95% CI 0.26 to 0.68). Lyons et al [37] reported a 68.3-m increase in 6-min timed walk (95% CI −106 to 243). Knight et al [40] reported mixed effects on VO_2 max for their sedentary time intervention (−2.71 mL/min/kg; 95% CI −7.05 to 1.63), their PA intervention (+2.06 mL/min/kg; 95% CI −3.20 to 7.32), and their combined intervention (+1.98 mL/min/kg; 95% CI −2.4 to 6.36).

Longer-Term Effects (6-12 Months)

No studies reported on this.

Behavioral Change Techniques

Of the 93 potential BCTs, only 31 were employed in the included studies. Studies included an average of 12 BCTs for intervention groups (range 5-21) and less than or equal to 2 BCTs in comparator groups (Multimedia Appendix 2).

The studies by Ashe et al [35] (PA), Lyons et al [37] (PA), and Silveira et al [39] (fitness, individual app) appeared to have the most effective interventions. Both the studies by Ashe et al and Lyons et al utilized apps that were linked to smart activity trackers. Furthermore, frequently employed BCTs in these effective interventions were *goal setting* (100%), *self-monitoring* (80%), *instructions on how to perform the behavior* (80%), *social reward* (80%), *social support* (60%), and *risk communication* (60%).

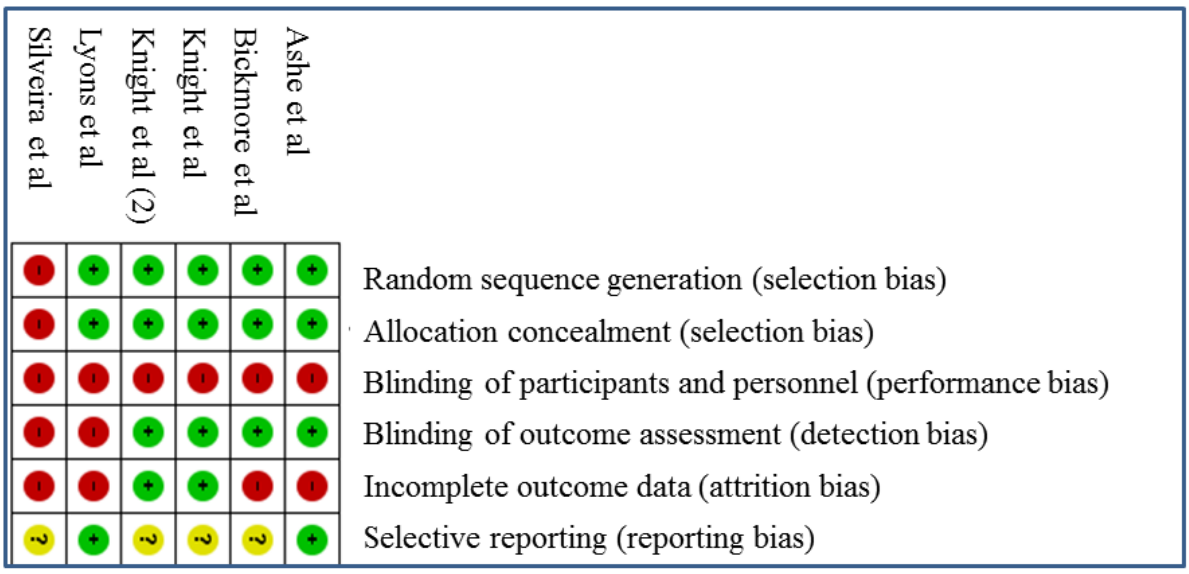
Discussion

Principal Findings

mHealth app interventions may be associated with decreases in sedentary time (SMD=−0.49), increases in PA (506 steps/day), and increases in fitness (SMD=0.31) in trials 3 months or shorter and with increases in PA (753 steps/day) in trials 6 months or longer. Results for all individual outcomes revealed trends in the same direction, but all results were inconclusive as the confidence intervals included zero.

Overall, risk of bias was low for all studies apart from the one by Silveira et al that was a nonrandomized trial (Multimedia Appendix 3, Figure 3), the results of which may have been subject to both selection and detection bias. The majority of studies had an RCT design (5/6 studies) and described random sequence generation (5/6 studies), allocation sequence generation (5/6 studies), and blinded outcome assessment (4/6 studies). All studies were judged at high risk of performance bias as it was not possible to blind participants to an mHealth app intervention. Generally, the study attrition rates were low. In addition, 2 studies were registered in a clinical trial registry [35,37], but the remaining 4 studies did not publish a protocol or register their trial [36,38-40].

Figure 3. Risk of bias summary.



There are a number of factors that contributed to heterogeneity. These include variation in comparator groups (active and nonactive), variation in the intervention package, the length of follow-up, and variation in the outcome measurements.

For PA, the statistical heterogeneity for the pooled estimate was high (I^2=80.5%). We tried to minimize heterogeneity by including the most homogenous measures (steps/day), but we identified a number of other sources of heterogeneity that could account for the I^2 value. The effect estimates across the studies were varied (3787 steps/day in the study by Ashe et al [35], 1607 steps/day in the study by Lyons et al [37], and 32 steps/day in the study by Bickmore et al [36]), and the greater success of the interventions by Ashe et al and Lyons et al may have been due to a number of factors.

Although users in all 3 studies synced the app to wearable devices, apps that were synced to smart activity trackers rather

than pedometers appeared to be more effective (Bickmore et al included a simple pedometer, whereas Ashe et al and Lyons et al used smart activity trackers). Furthermore, intervention packages with apps and some professional input seemed to be more effective than without (the study by Ashe et al included face-to-face group educational sessions, that by Lyons et al included weekly telephone counseling, and that by Bickmore et al had no health professional–led components). The nature of the comparator groups also appeared to influence the results. Participants in the comparator group in the study by Ashe et al had group education sessions not related to PA; in the study by Lyons et al, they underwent no intervention, whereas in the study by Bickmore et al, they were issued with pedometers. Finally, the studies by Ashe et al and Lyons et al had 3-month interventions, whereas the study by Bickmore et al had only 2-month interventions. Altogether, this may have led to Ashe et al and Lyons et al reporting larger effect sizes than Bickmore et al and hence contributed to statistical heterogeneity.

For sedentary time, there was little heterogeneity between the studies by Ashe et al and Lyons et al (I^2=0%), with both studies including smart wearable devices in their interventions and using nonactive comparator groups.

For physical fitness, statistical heterogeneity was zero. The pooled estimate utilized 3 different measures of fitness (6-min timed walk, VO_2 max, and fastest gait speed), although we used a standardized estimate to minimize the effect of this.

The overall GRADE certainty assessment of evidence for PA and sedentary time was moderate (Multimedia Appendix 4) and for physical fitness was low. The low certainty estimate for fitness was because the design of 1 of the 3 included trials was nonrandomized, the risk of bias for this same trial was high and risk of imprecision because of small sample size was high across all 3 studies.

Only 2 studies across this review had findings that reached statistical significance [35,39]. Common BCTs to both studies included goals and planning, feedback and monitoring, social support, and reward and threat. Both of these studies utilized wearable technologies that synced to the apps. Interestingly though, there are a number of emerging mHealth apps that use GPS technology on phones to track step count without needing wearable technology, but we found no studies testing their effectiveness in older adults.

Comparison With the Literature

A total of 3 reviews in the literature are relevant to this review. First, in their meta-analysis, Direito et al examined the effectiveness of mHealth interventions on PA and sedentary time in adults of all ages [41]. Second, in their systematic reviews, Muellman et al [28] and Jonkman et al [29] examined the effectiveness of eHealth interventions on just PA in older adults.

Direito et al [41] included 21 RCTs (n=700, objectively measured PA). Overall, 7 of these 21 interventions included mHealth apps (the rest included other mHealth), and only 1 of these studies included older adults (Knight et al [38]). They found that mHealth interventions led to decreased sedentary time compared with control (SMD=−0.26; 95% CI −0.53 to 0). Although in the same direction and a similar effect size to our result, their result was statistically significant. Direito et al [41] examined BCTs utilized across the interventions and found that a smaller range of BCTs was employed in interventions (mean 6.9, range 2-12) and a larger range in the control group (mean 3.1, range 0-10) in comparison with our findings. This may be because app technology (in comparison with all other mHealth technology) has the potential to offer more BCTs with minimal increase in time/cost. Only 33% of the included studies by Direito et al utilized apps compared with all of our included studies.

In their narrative systematic reviews, Muellman et al [28] and Jonkman et al [29] concluded that eHealth interventions (including internet-based, telephone-based, and text messaging–based interventions) can promote PA in the short term among adults older than 55 years, but evidence on long-term effects is still lacking. Our meta-analysis was inconclusive in supporting this finding, but we observed a trend of mHealth app interventions leading to increases in step count. Only 1 of the 4 studies [36] that we included in our review reporting PA outcomes was included in the reviews by Muellman et al and Jonkman et al. What could the trends we report here mean in the context of the average behavior levels of an older adult? In their review, Tudor-Locke et al [42] found that healthy older adults take an average of 2000 to 9000 steps/day, a very broad range reflecting the natural diversity of abilities common to older adults. In addition, they concluded that 30 min of daily MVPA accumulated in addition to habitual daily activities in healthy older adults is equivalent to taking approximately 7000 to 10,000 steps/day. We report that mHealth app interventions led to a trend of increasing step count (an extra 753 steps/day). If mHealth app interventions could provide an extra 753 steps/day over the longer term, this may represent over 10% (753/7000) of the step count, which Tudor-Locke et al equate to required daily PA levels. This could be a small but potentially clinically significant change at the population level. Berkemeyer et al [43] reported that men and women aged 60 to 70 years spent 25 min/day and 19 min/day, respectively, doing MVPA (>2020 counts per minute threshold). In this context, the mean increase of 22 min/day in MVPA following an mHealth app intervention reported by Ashe et al [35] would be enough to get individuals to meet activity guidelines. However, as only 1 study reported MVPA as an outcome and the 2 other studies reporting PA reported an inconclusive effect, more studies are needed to reach firm conclusions on whether mHealth app interventions may lead to clinically significant increases in PA levels.

Strengths and Limitations

The current meta-analysis is the first to assess the effectiveness of mHealth app interventions on sedentary time, PA, and fitness in older adults. We report here a reproducible and strong review of the current evidence, having published a prospective protocol with PROSPERO, utilized Cochrane and GRADE protocols, and undertaken a comprehensive search. We highlight the limited the number of primary studies and sample sizes in this area of research. We sought to describe sources of heterogeneity across intervention packages, comparator groups (active and

nonactive), duration of intervention, and length of follow-up in detail.

The included populations limit generalizability of these results. We set a low cutoff point for our definition of older adults, which may limit generalizability to *older* adults. Only 2 studies included adults older than 79 years. All interventions were delivered in high-income countries, with only 1 European country and no countries from Asia or Africa. The focus of this review was community-dwelling older adults. Given the nature of the intervention, all of the included studies excluded individuals with severe illnesses or disease limiting the ability to walk. Therefore, this review primarily represents a healthier older population. In addition, the studies by Ashe et al, Bickmore et al, and Lyons et al had inclusion criteria targeting particularly inactive adults (eg, <60 min PA/week or no PA usually). Inactive adults may have stronger habits to initially change but may have greater opportunity for change.

Conclusions and Implications

Mobile app interventions may be effective in decreasing sedentary time, increasing PA, and increasing fitness in trials 3 months or shorter and increasing PA in trials 6 months or longer, but we cannot conclude changes with certainty. Features that appeared to be common to effective app interventions included syncing to smart activity monitors; employing BCTs such as goal setting, self-monitoring, instructions on how to perform the behavior, and social reward; and combining apps with professional support. We found that the effectiveness of mHealth app interventions on sedentary time, PA, and fitness has been evaluated in very few studies. Furthermore, those studies that exist have small sample sizes. This review indicates the need for larger, robust RCTs into mHealth app interventions in older adults to power a future meta-analysis to reach firm conclusions on the effectiveness of mHealth app interventions in producing sustained clinical important changes in sedentary time, PA, and fitness levels.

We need to clarify if apps are associated with behavior change in the short term and, more importantly, the degree to which the changes are sustained.

With increasingly time-pressured health care systems, mHealth app interventions that can be tailored, yet delivered fast and cheaply, are potentially useful. Although technology-based interventions are becoming more commonplace in the general adult population, there is a need for a stronger evidence base to underpin interventions and for targeting older adults. The opportunity to utilize mHealth app interventions in older adults, significant numbers of whom now carry smartphones, should not be missed. We report that mHealth app interventions are a relatively underexplored tool to change PA, sedentary time, and physical fitness in older adults and recommend more attention to their utilization.

Acknowledgments

DAY was funded by a National Institute for Health Research (NIHR) doctoral fellowship (DRF-2017-10-121). SH was supported by the Lifelong Health and Wellbeing cross-council program, the Medical Research Council (MC_UU_12015/4), and Canadian Institutes of Health Research (FRN 146766). SJG is an NIHR senior investigator. The University of Cambridge has received salary support in respect of SJG from the National Health Service in the East of England through the Clinical Academic Reserve.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Literature search strategy: MEDLINE.
[DOCX File , 17 KB - jmir_v21i11e14343_app1.docx]

Multimedia Appendix 2
Summary of BCTs.
[DOCX File , 25 KB - jmir_v21i11e14343_app2.docx]

Multimedia Appendix 3
Risk of Bias Descriptions.
[DOCX File , 24 KB - jmir_v21i11e14343_app3.docx]

Multimedia Appendix 4
GRADE Certainty Assessment.
[DOCX File , 16 KB - jmir_v21i11e14343_app4.docx]

References

1. Harvey JA, Chastin SF, Skelton DA. Prevalence of sedentary behavior in older adults: a systematic review. Int J Environ Res Public Health 2013 Dec 2;10(12):6645-6661 [FREE Full text] [doi: 10.3390/ijerph10126645] [Medline: 24317382]

2. Taylor D. Physical activity is medicine for older adults. Postgrad Med J 2014 Jan;90(1059):26-32 [FREE Full text] [doi: 10.1136/postgradmedj-2012-131366] [Medline: 24255119]
3. Samani N. British Heart Foundation. Impact 2017 Aug;2017(6):40-41. [doi: 10.21820/23987073.2017.6.40]
4. Arem H, Moore SC, Patel A, Hartge P, de Gonzalez AB, Visvanathan K, et al. Leisure time physical activity and mortality: a detailed pooled analysis of the dose-response relationship. JAMA Intern Med 2015 Jun;175(6):959-967 [FREE Full text] [doi: 10.1001/jamainternmed.2015.0533] [Medline: 25844730]
5. Blair SN, Morris JN. Healthy hearts--and the universal benefits of being physically active: physical activity and health. Ann Epidemiol 2009 Apr;19(4):253-256. [doi: 10.1016/j.annepidem.2009.01.019] [Medline: 19344864]
6. Fogelholm M. Physical activity, fitness and fatness: relations to mortality, morbidity and disease risk factors. A systematic review. Obes Rev 2010 Mar;11(3):202-221. [doi: 10.1111/j.1467-789X.2009.00653.x] [Medline: 19744231]
7. Sui X, LaMonte MJ, Laditka JN, Hardin JW, Chase N, Hooker SP, et al. Cardiorespiratory fitness and adiposity as mortality predictors in older adults. J Am Med Assoc 2007 Dec 5;298(21):2507-2516 [FREE Full text] [doi: 10.1001/jama.298.21.2507] [Medline: 18056904]
8. Kodama S, Saito K, Tanaka S, Maki M, Yachi Y, Asumi M, et al. Cardiorespiratory fitness as a quantitative predictor of all-cause mortality and cardiovascular events in healthy men and women: a meta-analysis. J Am Med Assoc 2009 May 20;301(19):2024-2035. [doi: 10.1001/jama.2009.681] [Medline: 19454641]
9. Harber MP, Kaminsky LA, Arena R, Blair SN, Franklin BA, Myers J, et al. Impact of cardiorespiratory fitness on all-cause and disease-specific mortality: advances since 2009. Prog Cardiovasc Dis 2017;60(1):11-20. [doi: 10.1016/j.pcad.2017.03.001] [Medline: 28286137]
10. Al-Mallah MH, Sakr S, Al-Qunaibet A. Cardiorespiratory fitness and cardiovascular disease prevention: an update. Curr Atheroscler Rep 2018 Jan 16;20(1):1. [doi: 10.1007/s11883-018-0711-4] [Medline: 29340805]
11. Halaweh H, Willén C, Svantesson U. Association between physical activity and physical functioning in community-dwelling older adults. Eur J Physiother 2017;19(1):40-47. [doi: 10.1080/21679169.2016.1240831]
12. Lee J, Chang RW, Ehrlich-Jones L, Kwoh CK, Nevitt M, Semanik PA, et al. Sedentary behavior and physical function: objective evidence from the Osteoarthritis Initiative. Arthritis Care Res (Hoboken) 2015 Mar;67(3):366-373 [FREE Full text] [doi: 10.1002/acr.22432] [Medline: 25155652]
13. Guure C, Ibrahim N, Adam M, Said S. Impact of physical activity on cognitive decline, dementia, and its subtypes: meta-analysis of prospective studies. Biomed Res Int 2017;2017:9016924 [FREE Full text] [doi: 10.1155/2017/9016924] [Medline: 28271072]
14. Steffl M, Bohannon RW, Sontakova L, Tufano JJ, Shiells K, Holmerova I. Relationship between sarcopenia and physical activity in older people: a systematic review and meta-analysis. Clin Interv Aging 2017;12:835-845 [FREE Full text] [doi: 10.2147/CIA.S132940] [Medline: 28553092]
15. Biswas A, Oh PI, Faulkner GE, Bajaj RR, Silver MA, Mitchell MS, et al. Sedentary time and its association with risk for disease incidence, mortality, and hospitalization in adults: a systematic review and meta-analysis. Ann Intern Med 2015 Jan 20;162(2):123-132. [doi: 10.7326/M14-1651] [Medline: 25599350]
16. Baxter S, Blank L, Johnson M, Everson-Hock E, Woods HB, Goyder E, et al. Interventions to Promote or Maintain Physical Activity during and after the Transition to Retirement: An Evidence Synthesis. Southampton, UK: NIHR Journals Library; 2016.
17. Hobbs N, Godfrey A, Lara J, Errington L, Meyer TD, Rochester L, et al. Are behavioral interventions effective in increasing physical activity at 12 to 36 months in adults aged 55 to 70 years? A systematic review and meta-analysis. BMC Med 2013 Mar 19;11:75 [FREE Full text] [doi: 10.1186/1741-7015-11-75] [Medline: 23506544]
18. Das P, Horton R. Physical activity-time to take it seriously and regularly. Lancet 2016 Sep 24;388(10051):1254-1255. [doi: 10.1016/S0140-6736(16)31070-4] [Medline: 27475269]
19. Gardner B, Smith L, Lorencatto F, Hamer M, Biddle SJ. How to reduce sitting time? A review of behaviour change strategies used in sedentary behaviour reduction interventions among adults. Health Psychol Rev 2016;10(1):89-112 [FREE Full text] [doi: 10.1080/17437199.2015.1082146] [Medline: 26315814]
20. Martin A, Fitzsimons C, Jepson R, Saunders DH, van der Ploeg HP, Teixeira PJ, EuroFIT consortium. Interventions with potential to reduce sedentary time in adults: systematic review and meta-analysis. Br J Sports Med 2015 Aug;49(16):1056-1063. [doi: 10.1136/bjsports-2014-094524] [Medline: 25907181]
21. Duke SciPol. 2015. Mobile Medical Applications: Guidance for Industry and Food and Drug Administration Staff URL: http://www.scipol.duke.edu/content/mobile-medical-applications-guidance-industry-and-food-and-drug-administration-staff-final [accessed 2019-10-28]
22. Sjöström M, Lindholm L, Samuelsson E. Mobile app for treatment of stress urinary incontinence: a cost-effectiveness analysis. J Med Internet Res 2017 May 8;19(5):e154 [FREE Full text] [doi: 10.2196/jmir.7383] [Medline: 28483745]
23. Research2Guidance. 2017. MHealth App Economics Report 2017/2018 URL: https://research2guidance.com/product/connectivity-in-digital-health/ [accessed 2019-10-28]
24. Smith A. Pew Research Center. 2014 Apr 3. Older Adults and Technology Use URL: http://www.pewinternet.org/2014/04/03/older-adults-and-technology-use/ [accessed 2019-05-31]

25. Ofcom. 2017 Jun 14. Rise of the Social Seniors revealed URL: https://www.ofcom.org.uk/about-ofcom/latest/media/media-releases/2017/rise-social-seniors#4 [accessed 2019-05-31]
26. Deloitte. 2018. Global Mobile Consumer Survey 2017: The UK Cut URL: http://www.deloitte.co.uk/mobileuk/?id=gb:2el:3ms:4mobconsumer:5awa:6ci:prmobconsumer&_ga=2.149261046.1700363234.1523370865-1476812904.1523370865#smartphone-penetration-by-age-group-between [accessed 2019-05-31]
27. National Health Services. 2018. NHS Apps Library URL: https://apps.beta.nhs.uk/ [accessed 2018-05-31]
28. Muellmann S, Forberger S, Möllers T, Bröring E, Zeeb H, Pischke CR. Effectiveness of eHealth interventions for the promotion of physical activity in older adults: a systematic review. Prev Med 2018 Mar;108:93-110. [doi: 10.1016/j.ypmed.2017.12.026] [Medline: 29289643]
29. Jonkman NH, van Schooten KS, Maier AB, Pijnappels M. eHealth interventions to promote objectively measured physical activity in community-dwelling older people. Maturitas 2018 Jul;113:32-39. [doi: 10.1016/j.maturitas.2018.04.010] [Medline: 29903646]
30. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
31. Cochrane PH. CPHG Data Extraction Template. Cochrane Website 2011:1.
32. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
33. Higgins JP, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, Cochrane Bias Methods Group, Cochrane Statistical Methods Group. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. Br Med J 2011 Oct 18;343:d5928 [FREE Full text] [doi: 10.1136/bmj.d5928] [Medline: 22008217]
34. Cochrane Training. GRADE Handbook URL: https://training.cochrane.org/resource/grade-handbook [accessed 2019-05-31]
35. Ashe MC, Winters M, Hoppmann CA, Dawes MG, Gardiner PA, Giangregorio LM, et al. 'Not just another walking program': Everyday Activity Supports You (EASY) model-a randomized pilot study for a parallel randomized controlled trial. Pilot Feasibility Stud 2015;1:4 [FREE Full text] [doi: 10.1186/2055-5784-1-4] [Medline: 27175291]
36. Bickmore TW, Silliman RA, Nelson K, Cheng DM, Winter M, Henault L, et al. A randomized controlled trial of an automated exercise coach for older adults. J Am Geriatr Soc 2013 Oct;61(10):1676-1683. [doi: 10.1111/jgs.12449] [Medline: 24001030]
37. Lyons EJ, Swartz MC, Lewis ZH, Martinez E, Jennings K. Feasibility and acceptability of a wearable technology physical activity intervention with telephone counseling for mid-aged and older adults: a randomized controlled pilot trial. JMIR Mhealth Uhealth 2017 Mar 6;5(3):e28 [FREE Full text] [doi: 10.2196/mhealth.6967] [Medline: 28264796]
38. Knight E, Stuckey MI, Petrella RJ. Health promotion through primary care: enhancing self-management with activity prescription and mHealth. Phys Sportsmed 2014 Sep;42(3):90-99. [doi: 10.3810/psm.2014.09.2080] [Medline: 25295771]
39. Silveira P, van de Langenberg R, van Het Reve E, Daniel F, Casati F, de Bruin ED. Tablet-based strength-balance training to motivate and improve adherence to exercise in independently living older people: a phase II preclinical exploratory trial. J Med Internet Res 2013 Aug 12;15(8):e159 [FREE Full text] [doi: 10.2196/jmir.2579] [Medline: 23939401]
40. Knight E, Stuckey MI, Petrella RJ. Prescribing physical activity through primary care: does activity intensity matter? Phys Sportsmed 2014 Sep;42(3):78-89. [doi: 10.3810/psm.2014.09.2079] [Medline: 25295770]
41. Direito A, Carraça E, Rawstorn J, Whittaker R, Maddison R. mHealth technologies to influence physical activity and sedentary behaviors: behavior change techniques, systematic review and meta-analysis of randomized controlled trials. Ann Behav Med 2017 Apr;51(2):226-239. [doi: 10.1007/s12160-016-9846-0] [Medline: 27757789]
42. Tudor-Locke C, Craig CL, Aoyagi Y, Bell RC, Croteau KA, de Bourdeaudhuij I, et al. How many steps/day are enough? For older adults and special populations. Int J Behav Nutr Phys Act 2011 Jul 28;8:80 [FREE Full text] [doi: 10.1186/1479-5868-8-80] [Medline: 21798044]
43. Berkemeyer K, Wijndaele K, White T, Cooper AJ, Luben R, Westgate K, et al. The descriptive epidemiology of accelerometer-measured physical activity in older adults. Int J Behav Nutr Phys Act 2016 Jan 7;13:2 [FREE Full text] [doi: 10.1186/s12966-015-0316-z] [Medline: 26739758]

Abbreviations

BCT: behavioral change technique
eHealth: electronic health
GRADE: Grading of Recommendations, Assessment, Development, and Evaluation
mHealth: mobile health
MVPA: moderate- to vigorous-intensity physical activity
NIHR: National Institute for Health Research
PA: physical activity
PROSPERO: Prospective Register of Systematic Reviews

RCT: randomized controlled trial
SMD: standardized mean difference
VO2 max: maximal oxygen uptake

Edited by G Eysenbach; submitted 17.04.19; peer-reviewed by G Signorelli, M Hart, KL Mauco; comments to author 19.06.19; revised version received 01.08.19; accepted 24.09.19; published 28.11.19.

Please cite as:
Yerrakalva D, Yerrakalva D, Hajna S, Griffin S
Effects of Mobile Health App Interventions on Sedentary Time, Physical Activity, and Fitness in Older Adults: Systematic Review and Meta-Analysis
J Med Internet Res 2019;21(11):e14343
URL: http://www.jmir.org/2019/11/e14343/
doi:10.2196/14343
PMID:31778121

Original Paper

Exploring People's Candidacy for Mobile Health–Supported HIV Testing and Care Services in Rural KwaZulu-Natal, South Africa: Qualitative Study

Oluwafemi Adeagbo[1,2,3], PhD; Carina Herbst[1], MSc; Ann Blandford[2], PhD; Rachel McKendry[2], PhD; Claudia Estcourt[4], PhD; Janet Seeley[1,5], PhD; Maryam Shahmanesh[1,2], PhD

[1]Africa Health Research Institute, KwaZulu-Natal, Mtubatuba, South Africa

[2]University College London, London, United Kingdom

[3]University of Johannesburg, Auckland Park, Johannesburg, South Africa

[4]Glasgow Caledonian University, London, United Kingdom

[5]London School of Hygiene and Tropical Medicine, London, United Kingdom

Corresponding Author:
Oluwafemi Adeagbo, PhD
Africa Health Research Institute
KwaZulu-Natal
R168 Hlabisa Road Somkhele
PO Box 198
Mtubatuba, 3935
South Africa
Phone: 27 355507695
Email: Oluwafemi.Adeagbo@ahri.org

Abstract

Background: The use of mobile communication technologies (mHealth: mobile health) in chronic disease management has grown significantly over the years. mHealth interventions have the potential to decentralize access to health care and make it convenient, particularly in resource-constrained settings. It is against this backdrop that we aimed to codevelop (with potential users) a new generation of mobile phone–connected HIV diagnostic tests and Web-based clinical care pathways needed for optimal delivery of decentralized HIV testing, prevention, and care in low- and middle-income countries.

Objective: The aim of this study was to understand ways in which an mHealth intervention could be developed to overcome barriers to existing HIV testing and care services and promote HIV self-testing and linkage to prevention and care in a poor, HIV hyperendemic community in rural KwaZulu-Natal, South Africa.

Methods: A total of 54 in-depth interviews and 9 focus group discussions were conducted with potential users (including health care providers) in 2 different communities. Theoretically informed by the candidacy framework, themes were identified from the interview transcripts, manually coded, and thematically analyzed.

Results: Participants reported barriers, such as fear of HIV identity, stigma, long waiting hours, clinic space, and health care workers' attitudes, as major impediments to effective uptake of HIV testing and care services. People continued to reassess their candidacy for HIV testing and care services on the basis of their experiences and how they or others were treated within the health systems. Despite the few concerns raised about new technology, mobile phone–linked HIV testing was broadly acceptable to potential users (particularly men and young people) and providers because of its privacy (individual control of HIV testing over health provider–initiated testing), convenience (individual time and place of choice for HIV testing versus clinic-based testing), and time saving.

Conclusions: Mobile phone–connected HIV testing and Web-based clinical care and prevention pathways have the potential to support access to HIV prevention and care, particularly for young people and men. Although mHealth provides a way for individuals to test their candidacy for HIV services, the barriers that can make the service unattractive at the clinic level will also need to be addressed if potential demand is to turn into actual demand.

(J Med Internet Res 2019;21(11):e15681) doi:10.2196/15681

KEYWORDS

mHealth; antiretroviral therapy; HIV testing; South Africa; candidacy framework

Introduction

HIV Burden in South Africa and KwaZulu-Natal Province

The United Nations' Sustainable Development Goals herald a major commitment to accelerate the pace of progress made in tackling the HIV epidemic [1]. However, despite huge advances in antiretroviral therapies (ARTs) and antiretroviral-based prevention, many countries are struggling to reach this target, as they do not have the tools or mechanisms to deliver ART at scale to people in need [2,3]. Globally, South Africa (SA) has the highest number of people living with HIV (People living with HIV: estimated 7.9 million), with 20% of the total global antiretrovirals use and an HIV prevalence rate of 14% in 2017 [4,5]. With its large HIV prevention programs, SA has made significant progress in reducing AIDS-related mortality and the number of new HIV infections. However, annual new HIV infections (231,000 in 2017) are still high [4].

The province of KwaZulu-Natal (KZN) is disproportionately affected by the epidemic, with an estimated 18.1% prevalence rate [4], whereas in the uMkhanyakude district in the north of the province, the site of the research reported in this paper, it is estimated at 30% in the general population, with 5% to 7% annual HIV incidence rate in young women and men [6]. Furthermore, there is a high HIV-related mortality rate in men and less than 50% annual HIV testing uptake among men under 29 years, and only 50% of those diagnosed in the general population reach clinical services within 1 year. At the same time, clinic attendances for treatment have risen by 300% over the past 6 years, placing huge pressures on health systems, with escalating costs [6,7]. Therefore, there is a need to develop interventions that can close the gap in HIV testing, prevention, and treatment, to reduce HIV incidence and mortality, while reducing the current pressures on primary health care clinics. In this study, we sought to understand the current barriers and facilitators to HIV testing and care services to codevelop with potential users a mobile health (mHealth) intervention that will aid people's engagement with HIV care services. We used the *candidacy framework*, described below, to structure our findings. Moreover, the proposed mHealth intervention is described below.

The Limit of Universal HIV Test and Treat and HIV Prevention Technologies

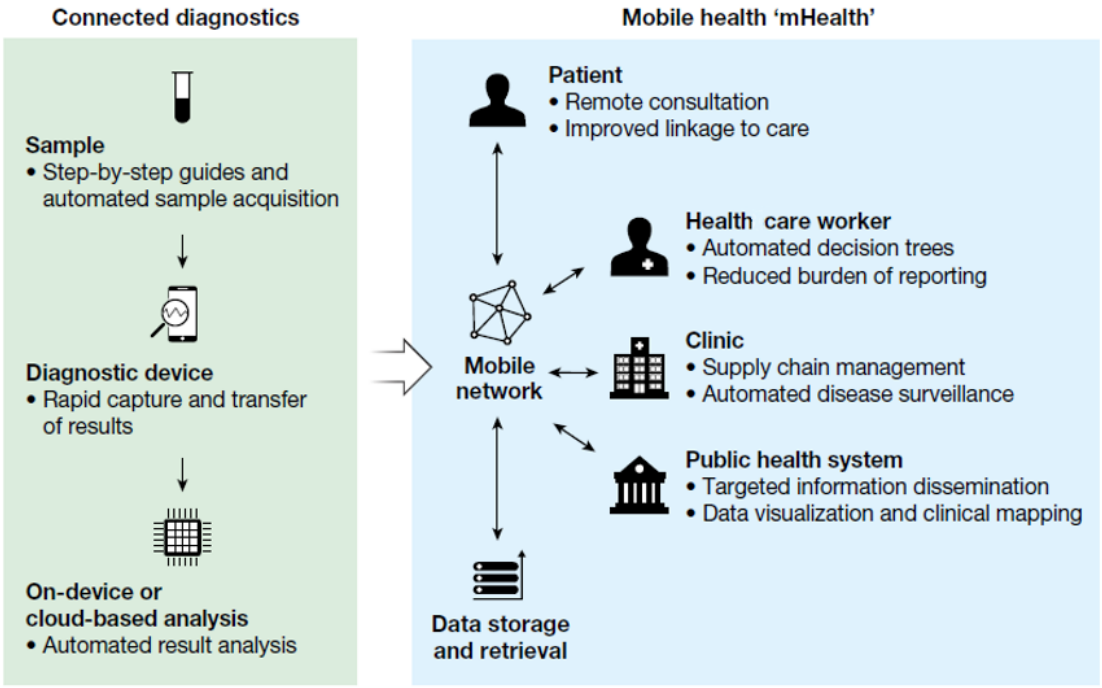
Universal HIV test and treat (UTT) and newer HIV prevention technologies, such as Pre-Exposure Prophylaxis (PrEP), have the potential to be game changers, but all require people to know their HIV status and engage in the cascade of care and prevention [7-10]. A treatment as prevention (TasP) trial conducted in rural SA failed to show an effect, partly because of social and structural barriers to link men and young people to care [7,11]. Similarly, a *test and start* study in Eswatini found that men still delay HIV testing and treatment until they have advanced disease; this is because of social barriers and stigma associated with accessing HIV care [12]. Although there was an effect on HIV incidence, the HPTN 071 trial (POPART) shows limited linkage to care for young men and women [13]. Despite the potential of UTT to halt HIV transmission, balancing social realities and science is very important to understand the context of HIV prevention across different settings [12,14]. Although stigma and other factors have been a major barrier to HIV care [12], research has shown that various HIV prevention technologies, such as HIV self-testing (HIVST), have the potential to widen access to testing, especially for hard-to-reach groups, such as adolescents and men [15-17].

However, for the game-changing potential of new innovations to be achieved in high-burden settings, such as SA, there is a need for those who test positive to link into immediate ART, as well as those who test negative to access combinations of HIV prevention interventions tailored to keep them negative, including but not limited to condoms, PrEP, and voluntary male circumcision. Therefore, there is an urgent need to identify cost-effective and acceptable technologies for population-based HIV testing, including self-testing with linkage and retention in care.

The Potential of Mobile Health in Closing the Gaps in the HIV Treatment Cascade

The use of mobile communication technologies (mHealth) in chronic disease management has grown significantly over the years. mHealth interventions have the potential to decentralize access to health care and make it convenient, particularly in resource-constrained settings [18-22]. It is well documented that mHealth interventions, such as SMS text messaging, have been used to support management of diseases, such as hypertension and diabetes, and to support HIV treatment [20,23,24]. In fact, the World Health Organization recommended the use of mobile phone–technologies for ART adherence and other chronic disease management [25]. With 76 million mobile phone subscribers, SA has the highest penetration rate in Africa, therefore it has potential for mHealth interventions [26,27]. Current innovations in HIVST, task shifting to community caregivers, and the rise in HIV treatment adherence clubs, try to support chronic disease management and reduce the pressure on health facilities by shifting the focus of HIV testing, treatment, and prevention away from health care facilities to the community. However, in the absence of robust and safe clinical pathways to support these interventions, their reach will be limited. There is now an opportunity with the growth in mobile technologies to provide digital support for these innovations (see Figure 1), while pushing the boundaries further to potentially supporting the entire continuum of HIV care, from primary prevention to long-term treatment and improved sexual health of people living with HIV.

Figure 1. Mobile health–connected diagnostics by Wood et al, 2019.



SA has a progressive mHealth strategy [27] to deliver improved HIV care services, but this emerging field is still in its infancy, with currently no safe and acceptable Web-based clinical care pathways and very limited knowledge of the feasibility and acceptability of implementing these technologies within existing care pathways. mHealth interventions are complex interventions that need to involve prospective users and service providers (SPs) in their development—considering the patterns and preferences for technology—to maximize uptake and scalability of interventions using mobile electronic devices, such as mobile phones, for the delivery of health care services and avoid inadvertently worsening health inequity [19,23,28]. The mAfrica study aimed to harness this growth in mobile phone usage and potential for enhanced diagnostic tests using inbuilt phone sensors (eg, camera) to interpret and display results and then direct people to local clinics to support virtual follow-up appointments and potentially provide access to antiretrovirals for both treatment and prevention in the community. In summary, this study aimed to codevelop (with potential users) a new generation of mobile phone–connected HIV diagnostic tests and Web-based clinical care pathways needed for optimal delivery of decentralized HIV testing, prevention, and care in low- and middle-income countries. This paper presents the qualitative data from the formative phase of the study to gain insight into potential users' perceptions of barriers to HIV testing and treatment, as well as their willingness to use the proposed app to promote HIVST and further linkage to care through Web-based clinical care pathways in rural KZN.

Methods

Overview

The research was conducted with 2 different communities in a rural subdistrict (228,000 population) of uMkhanyakude district in KZN. The 2 communities were different in terms of their development, access to basic amenities, and housing structures. The first community (CA) is a township that is more developed, with more amenities, population, and modern housing structures, compared with the second community (CB) that is rural in nature, with few basic amenities and housing structures. Each community has government-funded health clinics. The research was conducted by a group of social scientists (males and females) trained in qualitative data collection methods and competent in both IsiZulu (local language) and English. A total of 4 social scientists conducted the community in-depth interviews (IDIs) and focus group discussions (FGDs) in the local language, and a senior social scientist conducted the SPs' interviews in English. Participants were purposively sampled and recruited individually within the research communities to represent different age groups, gender, and geographical locations in the same district. In our experience, group discussions yield richer data around the normative thinking within demographic groups because of intergenerational and gender relationships if we divide groups by gender and age. This was particularly important, given our aims to explore the ways that different groups may respond differently to the intervention. Thus, community members were divided into 4 categories for IDIs and FGDs: young female, young male, older female, and older male. A total of 4 health care providers (2 female pharmacists and 2 male community HIV testing services providers) participated in IDIs, and 9 research nurses (3 males

and 6 females) participated in an FGD together; they were all regarded as SPs. Moreover, both community members and SPs (health care professionals) were regarded as *end users* in this study.

A total of 54 IDIs and 9 FGDs were conducted, with participants (both sexes aged 18-79 years) in both communities, between November 2017 and March 2018. Although our target was 5 IDIs for each group of community members, we slightly exceeded the numbers with some groups and stopped at saturation. The distribution across communities and categories is summarized in Table 1.

Table 1. Participant demographics and data collection method.

Population[a,b]	Age range (years) (IDIs[c])	IDIs conducted (n)	FGDs[d] conducted (n)	Age range (years) (FGDs)	FGD participants (n)
Community A (township)					
Young female	23-32	6	1	19-32	6
Young male	18-32	8	1	18-28	7
Older male	37-46	7	1	35-67	11
Older female	43-50	6	1	43-67	6
Community B (rural area)					
Young female	18-34	6	1	18-23	10
Young male	20-33	6	1	19-31	8
Older male	36-79	6	1	36-58	6
Older female	35-60	5	1	36-48	6
Community service providers (SPs) IDIs					
Pharmacist (F[e])	28 (SP1)	1	—[f]	—	—
Pharmacist (F)	34 (SP2)	1	—	—	—
HTS[g]/counsellor supervisor (M[h])	53 (SP3)	1	—	—	—
HTS counsellor (M)	30 (SP4)	1	—	—	—
Community service providers FGDs					
Research nurses: (F and M)	—	—	1	33-55	9 (6 F and 3 M)

[a]Data collection period: November 2017 to March 2018.

[b]General population: 18 to 79 years.

[c]IDI: in-depth interview.

[d]FGD: focus group discussion.

[e]F: female.

[f]Not applicable.

[g]HTS: HIV testing services

[h]M: male.

All IDIs and FGDs were conducted in participants' language and places of choice in the community after voluntary informed consent (written and verbal) was obtained. IDIs lasted between 30 and 60 min, whereas FGDs lasted between 90 and 120 min. Through semistructured interviews, we explored participants' views around barriers to HIV testing and treatment in their communities, as well as their willingness (including their fears and expectations) to use a digital technology to promote HIVST and further linkage to care. Moreover, the group discussions afforded us the opportunity to tease out the complexities, similarities, and differences in what participants said during individual interviews. Moreover, reflective summaries of all interviews conducted, as well as findings, were written down and discussed during several debriefing sessions with other team members. All interviews were digitally recorded, and those conducted in IsiZulu were transcribed verbatim and translated to English for thematic analysis. The data were stored and managed in a Web-based shared drive with limited access. All data were deidentified to protect confidentiality. The data were scrutinized and compared with the recordings by a senior social scientist (OA) and 3 senior research assistant supervisors for quality control. In addition, a 2-day *coding and analysis* workshop was held with the social scientists who conducted the fieldwork. During the workshop, emerging themes were identified, and coding was crosschecked for consistency across coders. Following a *candidacy framework* [29,30], as described below, themes were identified from the interview transcripts, manually coded, and thematically analyzed to present the subjective views of the participants. Key themes discussed in this study were agreed upon by the research team. The study

was approved by the Biomedical Research Ethics Committee of the University of KwaZulu-Natal, South Africa (Reference Number: BE435/17).

Theoretical Framework

The analysis for this paper is based on an adapted *candidacy framework* to explore community barriers and facilitators to HIV testing and treatment, as well as participants' willingness to use an mHealth technology [29-31]. The *candidacy framework* was originally constructed to investigate the process of accessing health care by vulnerable people in the United Kingdom [29], and it has been modified in different ways over the years, including to examine the utilization of health services by young people in KZN [32]. Taking a candidacy approach places the focus on the choices individuals make, as summarized in Table 2.

Individuals need to first recognize that they are eligible for a health care service and that people like them can access the service. Individuals presenting themselves at health care facilities affirm their candidacy for that service, but their eligibility is adjudicated by health care providers who decide their suitability. Once access has been established, continuing to use health care services requires navigating relevant systems (eg, time off work and transportation) and continuing to consider that the service meets their needs. The pathway to accessing care requires an active response from users and the ability to overcome the barriers [30,32,33]. We use the *candidacy framework* in our data analysis to show the complexities and intersections of multiple factors impeding and supporting individual access to HIV care services in a resource-constrained setting, as well as the ways mHealth-supported HIV testing can or cannot support users to explore their *candidacy* in relation to HIV prevention and care.

Table 2. Characteristics of the 7 stages of the candidacy framework by Mackenzie et al.

Stages of candidacy	Description of stages
Identification of candidacy	The process by which individuals come to view themselves as legitimate candidates for particular services
Navigation of services	Knowing how to interact with appropriate services in relation to identified candidacy
Permeability of services	Includes the level of explicit or implicit gatekeeping within a service and the complexity of its referral systems, referring to the *cultural alignment* between users and services
Appearing at services and asserting candidacy	The actions that individuals must take to assert their candidacy in an interaction with a health care professional
Adjudication by professionals	Candidacy, as expressed by service users, is validated or otherwise by health care professionals, which influences subsequent service offers
Offers of, resistance to, services	Emphasizes that follow-up services may be appropriately or inappropriately offered and that these may or may not be acted upon by service users
Operating conditions and local production of candidacy	This incorporates factors that influence decisions about subsequent service provision (eg, the resources available for addressing candidacy) and the kinds of contingent relationships that develop between professionals and service users over a few encounters

Results

The section is divided into 2 main parts: (1) barriers to the identification of candidacy for HIV testing and care and (2) assessing candidacy through HIVST and mobile phone–enabled linkage to care. The themes under these broad headings engage with key challenges faced by people in identifying their candidacy for HIV testing, treatment or prevention, and the potential of HIVST and mHealth technology to overcome the barriers of *identification of candidacy* by making it easier for people to assess their candidacy for access to HIV testing and care services.

Barriers to the Identification of Candidacy for HIV Testing and Linkage to Care

In this section, we focus on the barriers people face in recognizing themselves as *candidates* for HIV testing, treatment, or prevention. We highlight the complexities and multiple intersecting factors that shape people's everyday experiences of accessing HIV services in their locations.

Everybody Will Die One Day: Barriers to Individual Identification of Candidacy for HIV Testing

Our data show that it is difficult for participants, particularly men in both communities, to test for HIV. It seems some older women have transcended the challenges of HIV testing because of their engagement with sexual and reproductive health services. Men do not see themselves as candidates for HIV testing. They are more afraid of knowing their HIV status than HIV itself. This is partly because of the association of HIV with *assumed promiscuity* of men. Acquiring HIV is seen to be inevitable for men, and this is reflected in the below excerpt that everyone will die one day. The inevitability of HIV and the need to start HIV treatment one day are juxtaposed with the huge perceived social costs of knowing their HIV status (even if one identifies as a candidate), namely, discrimination and loss of loved ones, leading to delays in accessing HIV care services:

> *With men there is no specific reason except that it's a matter of reasoning...I would say they don't have a concrete reason as to why they don't want to. But they would say things like everybody has HIV or*

> *everybody will die one day, so am I, so why should I bother.* [IDI, M, 53, SP3]
>
> *They also don't want to know their status; they try to avoid discrimination. They're also afraid that they might lose their loved ones if they found out that they are HIV positive.* [IDI, Young Male, 26, CB]

HIV-related stigma makes it more complex for some men to identify and assert their candidacy for HIV testing, treatment, or prevention, as they anticipate navigating the available HIV services until their candidacy is adjudicated by health care professionals. The pathway of accessing health care is an iterative process that requires an active response from users, whereas the barriers, people's health needs, and perception of health services often create vulnerabilities [30,32]. For example, men (both young and old) are often seen as the major drivers of the HIV epidemic [34], and being seen at a local clinic, accessing HIV care, could reinforce suspicion, gossip, and stigmatization, even from local health care workers; therefore, men are reluctant to access HIV care services. As reported, some people (particularly men) are reluctant to test for HIV, as they will be judged to be HIV positive immediately when they are seen at their local clinic, even if just attending for testing. HIV-related stigma and unwillingness of men to utilize public clinic–based HIV testing and care services partly accounted for poor linkage to treatment in a TasP trial in KZN [7,11]. Our findings are supported by similar results from other sub-Saharan African studies on the impact of HIV-related stigma on people's access to HIV testing, prevention, and care [12,13,32,35-39].

Illness identity is a crucial phase of individuals' candidacy and multiple intersections that shape how it is navigated and adjudicated within the health care services. Previous studies have shown that people from disadvantaged settings are less likely to use preventive services and more likely to normalize *illnesses* because of their inability to see the positive side of health and their fear of being stigmatized or blamed by health care providers [30,33]. It was noted during the fieldwork that men are conscious of manliness being incompatible with illness; therefore, most of them living with HIV do not see themselves as candidates for HIV testing, and therefore do not know their status. Similar findings were reported about men across SA in a recent national HIV survey [4]. From the foregoing, perception of *HIV identity* and the blaming of men as drivers of the HIV epidemic make it more difficult for them to identify their candidacy for HIV testing and care, navigate the available HIV services, and appear at the clinic for their candidacy to be adjudicated by health care professionals. Our findings add to the knowledge and show that the fear of HIV-related stigma hinders people's access to HIV services, despite the apparent normalization of HIV as an inevitable (and invisible) epidemic among men.

They Can't Eat the Sweet While It's on Paper: Association of HIV With Socially Undesirable Behaviors

Some participants, particularly young people (males and females), are afraid to test for HIV because of the way in which this could mark them out as having engaged in socially undesirable behaviors, such as unprotected sex, alcohol and drug abuse to health care providers, the community, and even peers. As reported by participants, the metaphorical *sweet*, which represents unprotected sex, is seen as ubiquitous in males, despite the association with HIV acquisition. Participants' views are illustrated below:

> *From age 18 to age 28 [males and females], they are prone to be at risk of contracting HIV because you know they go to clubs and take alcohol and drugs...* [IDI, Male, 30, SP4]
>
> *Especially my friends...they say they can't eat the 'sweet while it's on paper' you see that thing, that's the thing that can get them infected.* [IDI, Young Male, 23, CA]
>
> *Oh! the amount of unprotected sex that goes on here it astonishing and young kids just don't understand; they don't understand the implications of their actions...* [IDI, Female, 34, SP2]

This association of HIV acquisition with behaviors that are considered socially unacceptable and irresponsible, such as unprotected sex and sex under the influence of alcohol and drugs, makes it difficult for people to identify themselves as candidates for HIV testing at their local clinics. Specifically, for young people, it would be hard for them to identify themselves as candidates for HIV testing, even if they had not engaged in these socially undesirable behaviors, because of the widespread association of HIV with stigmatized behaviors in the community. As discussed above, identification of one's candidacy for HIV testing is an important stage; however, participants continue to evaluate their candidacy on the basis of various factors, such as stigma. Research conducted elsewhere in rural SA further corroborates our findings that young people often delay seeking health care, particularly HIV testing, because of stigma [4,32,40].

That Right Turn Is the Problem: Complexities of Navigation, Appearing at Services, and Adjudication

The layout of their local clinic, long queues, and inconveniences related to accessing services were described as significant barriers for participants in both communities. Most participants complained about the time required to access care. For men, clinic structure and operational hours are not suitable, given that most men in the locality are breadwinners, and they need to work to sustain their families:

> *If you go for [HIV] testing and you left home early in the morning, you will see that people [clinics] are full, and you may end up not doing it...* [IDI, Older Male, 36, CB]

However, inadvertent disclosure of HIV status and the ways in which people are treated by health care providers were also raised as barriers to care:

> *Here in community A, the clinic is split into two and when you go to the HIV section you have to take a right turn. That right turn is the problem because even someone that is passing by in a taxi can see that you are going there and will therefore immediately assume that you are HIV positive.* [P3: FGD, Older Female, 43-67, CA]

> *At the clinic people are complaining about the attitudes. After you have tested, other are being shouted at and you end up think that everyone is looking at you. They call your name "hey! Come this side," people also fear that. And that a person would be seen by wrong people, I think this is the main challenge they face, it is difficult to go there.* [IDI, Young Female, 25, CB]

Our findings corroborate the findings of other studies that have described the structural barriers that hinder people's (males and females) utilization of HIV services [41-43]. However, these structural barriers were compounded by anticipation that those who want to access HIV services are easily identified and stigmatized if they are seen going to a certain direction in the clinic. Finally, the attitudes of some health care workers (nurses) were frowned at. Almost all the participants in both communities complained about the inappropriate attitudes of some of the nurses at their local clinics. Our data validate and add nuances to growing evidence that HIV testing and treatment in public clinics are hampered by lack of privacy, long waiting times, segregation of services, and discrimination by health care workers [7,44-46]. Put together, the barriers hindering access to HIV testing and care services revealed the intersectionality of multiple factors affecting the recognition of one's candidacy for HIV services, as well as health worker's attitude as a key component of the *adjudication* process.

When You Tell Them to Go to the Clinic, They Say They are Scared: Community-General Anxieties to Accessing HIV Testing and Care

The fear of an *HIV identity* was common in both communities, and this is reflected in participants' descriptions of *fear* of appearing at the clinic for HIV care services. Some participants are afraid of what people would say about them if they see them at the clinic accessing HIV care, whereas others fear finding out they are positive and the *unknown* life after. For example, a young woman said the following:

> *I don't want to stress myself, because once I know my [HIV] status I will now start to think that I am going to die...* [IDI, Young Female, 29, CA]

Particularly, men fear the questions they will be asked about the person they date, her HIV status, and number of their sexual partners. Moreover, some men are uncomfortable seeking health care from their local clinics that are often dominated by female clients and health care providers, as they may not be treated as candidates for HIV testing and care services. For example, a young man maintains the following:

> *People don't want to go to clinic because they [nurses] ask lot of questions because what I see is that they [mostly males] don't like to be asked question, they are scared of being asked questions, maybe a person [nurse] will ask you how many people have you dated and what was their status?* [IDI, Young Male, 21 CB]

The fear of being stigmatized at the local clinics is also a challenge, as HIV care sections are often segregated. A young man shared his sentiment during the FGD:

> *When you get to the clinic to test for HIV, the rooms for HIV testing are excluded from other rooms.*

These are small communities, and information circulates very fast; therefore, there are fears and anxieties of some people, particularly men and young people, about accessing HIV care services in their local communities. For some, knowing their HIV status seems like a death sentence; therefore, they refuse to test. As illustrated by the candidacy framework [30], the prediagnosis is an important phase of awareness or acceptance of one's candidacy to access HIV testing and care services, which could be marred by barriers, such as fear of unknown consequences. Although we sought to understand how to improve people's candidacy for HIV care, our findings correspond to the results of other studies conducted in similar environs and show that there is still a general anxiety when it comes to HIV testing and treatment, particularly among young people and men [7,38]. Similarly, an Eswatini study highlights that most men wait until their health deteriorates before they access HIV testing and care because of fear and stigma [12]. Given the local conditions and both internal and external stigma, the prediagnosis phase that was supposed to be the initial recognition of our participants' candidacy to access HIV testing and care services was characterized by fear of an *HIV identity*.

Assessing Candidacy Through HIV Self-Testing and Mobile Phone–Enabled Linkage to Care

We now focus on the potential of HIVST and mHealth technology to identify candidates for HIV testing and facilitate linkage to care. We demonstrated HIVST and described the hypothetical functions of the proposed mHealth app to participants in the 2 communities. Our descriptions entail how they would conduct HIVST and be supported via the app, using mobile phone sensors (eg, camera) to interpret and display test results and then direct people to local clinics, to support virtual follow-up appointments and potentially provide access to antiretrovirals for both treatment and prevention (eg, Figure 1). A key finding discussed below is that an mHealth intervention can provide a way for people to test their candidacy for HIV service, which may eventually help to overcome the first barrier in the candidacy cascade.

Potential of HIV Self-Testing to Identify Candidates for HIV Testing

Generally, participants valued *the potential privacy and convenience* associated with HIVST. Given that HIV-related stigma is widespread in both communities studied, HIVST is generally acceptable, particularly among men and young people, as it gives them control of HIV testing, privacy, and convenience, thereby attracting candidates who were previously deterred by clinic-based testing. Participants are generally enthusiastic about HIVST, as reflected in the following quote from a 23-year-old young man:

> *I think it [HIVST] gives you that braveness that you test yourself.*

A young woman also shared her sentiments about HIVST:

> *It [HIVST] will help me because, going to the clinic takes time, hence it will save time and money for transport...* [IDI, Young Female, 23, CB]

Our data show participants' high willingness to perform HIVST, and this may be a strategy to identify candidates and improve diagnosis by placing HIV testing and disclosure within individuals' control. Similarly, HIVST appears highly acceptable in other settings because of its convenience and privacy, despite concerns about counseling and linkage to care [16,20,47].

However, there is a general concern about social harms, such as suicide, depression, and alcoholism, which could come with HIVST. These concerns about HIVST are not peculiar to our study, as they have been documented in other studies [48,49]. A few participants raised concerns that people may commit suicide after testing HIV positive without health care provider support, but this was deemed less of an issue by providers, given the high level of community awareness about HIV management. Abusing alcohol was seen more as a way of coping with their *HIV+ identity*, as this is widely stigmatized in both communities, than a consequence of HIVST:

> *I don't think suicide is a problem now, I think alcoholism. It [HIV+ status] drives them to that now…that's how they try to deal with it. I don't think people will kill themselves I know a lot of people that are sick [HIV+], and they do get depressed.* [IDI, Female, 28, SP1]

Thus, HIVST may allow people (particularly men and young people) to explore their candidacy for HIV testing in private and overcome fears of being assumed to be positive even if one has a negative test in clinic, but it does not erase the fear of the consequences of an *HIV identity* and other external factors or stigma [50]. Despite people's concerns and anxieties around new HIVST technology, several studies and systematic reviews have shown high uptake of HIVST, as well as its potential to attract first-time testers, particularly hard-to-reach populations [16,20,47,51].

Potential of Mobile Health Intervention to Overcome the Barriers of Individual Candidacy for HIV Care Services

This theme explores the potential of a mobile phone–connected Web-based clinical care pathway to link people to HIV prevention and care. Mobile phone–enabled linkage to care was broadly acceptable to potential users (including providers) because of the ability to test their candidacy for HIV care in private and at their own convenience. Participants, particularly men and young people, believe that linking to HIV care services through their phones provides opportunities to save time and reduce experienced stigma. The privacy and convenience the proposed app may provide is reflected in the following excerpts:

> *I think it is a good idea that there is technology like this. And I think it will work mostly for young people, because youth are the people that mostly get infected with HIV...Yes, I think they will be motivated, because it entails the privacy. Hence, no one will see other people's results since people will be testing alone. Then this APP will link with the clinic, then the clinic will continue to help you...* [IDI, Young Female, 30, CA]

> *Because it will be easy by that time to know yourself and not be scared... because the time you link with the clinic you will be able to make an appointment to collect your pills, and if you come to the clinic because you have made an appointment you will just walk in and take your pills and go maybe that can help and prevent the loss of self-confidence I think it will work a lot and I know young people they like to get into things like that.* [IDI, Young Male, 26, CB]

To participants, the proposed app, coupled with HIVST, could be a game changer by providing a way for an individual to identify and test their *candidacy* for HIV services. The particular appeal of the intervention was that it could overcome the experience of stigma and structural barriers of clinic-based testing by providing privacy and convenience. However, it does not completely erase individuals' fear of an HIV identity (self-stigmatization), as well as the barriers of navigating and appearing at HIV services for further care. The barriers that make the services unattractive at fixed clinic level will need to be addressed if potential demand is to turn into actual demand. Moreover, the development, implementation, and evaluation of the app will provide data about its receptivity and usage, as technology *likeness* does not mean that an individual automatically identifies as a candidate for its use.

Despite the declared willingness of participants to use the technology, a few concerns were raised about digital literacy and data consumption. Some participants raised concerns that some people would be left out, as they cannot afford smartphones, whereas others (particularly old people) with a smartphone cannot navigate it effectively, which could impede the assertion of their candidacy for the service. The concern around literacy and linkage is reflected in the following excerpt from a male SP:

> *Yeah, my concern is that if people are not well informed about the service that is linking the client after the self-test, then there might be a problem...*

The concern on data is reflected in the following excerpts from an older male FGD participant:

> *Where can we find the money to access this app? The problem will be airtime.*

These echo concerns around how technology and the digital divide can increase rather than reduce health inequalities [20].

Most participants were enthusiastic and willing to use the proposed app that has the potential to provide access to real-time HIV care, interpreting test result, counseling, relevant health information, and further linkage. Our results correspond to the findings of a recent study conducted with men who have sex with men (MSM) in China, using WeChat (app) to identify and assert their candidacy for HIV testing, prevention, and care. That study found that MSM who performed HIVST in private (eg, at home) and tested positive sought support via the app in interpreting their HIV results and linkage to treatment [20]. Other studies have also shown that software apps have the potential to reach and encourage hard-to-reach populations to

assert their candidacy for a particular service [52,53]. Owing to the self-sufficient nature of HIVST in which individuals can test and interpret their HIV results in a private location without the support of a trained counselor, the proposed app will allow people to test whether they are candidates for HIV care, and it may subsequently facilitate HIV testing and linkage to treatment if users see themselves as candidates for the service.

Partnership for Change: Community Cocreation of Mobile Health Intervention to Improve Access to HIV Care Services

To cocreate a user-friendly mHealth technology that could be used at home and in clinic settings to overcome some of the barriers identified above, end users and health care providers were asked about additional features they would like to see in the proposed app and whether they are happy to receive health-related messages on their mobile phones. Generally, participants were enthusiastic to receive health messages on their personal mobile phones and happy to describe features they believed would support them assess their candidacy for testing and care. In keeping with other studies, they found the concept empowering, convenient, and with the potential to reduce the experience of stigma [20,54,55]. Key features were therefore to reduce the time spent at the clinic and support privacy and confidentiality, as reflected in these excerpts:

> *It's good because sometimes I need to go to the clinic, yet I don't go because I am scared of all the people. If I get a message on my phone notifying me of my parcel's [HIV medication] readiness, I can go to the clinic knowing that I will be in and out quickly.* [P2: FGD, Older Female, 36-48, CB]

> *I think the counselling one must be there because counselling make person to be positive about that thing…That's why the App must be locked, you will go there and put your personal details so the other person can't access your information because that will mean your information is open to everyone if it will be just the password only, there must also be other things like questions that will be asked like a confirm questions. It can ask for your mother's name or your relative...* [P1: FGD, Young Male, 18-28, CA]

Participants in our study tried to limit physical contact with health facilities; therefore, they were interested in the alternative method that is private and confidential and that would help in asserting their candidacy by easing their access to HIV testing and care services. In parallel though, participants wanted counseling and support for some people to be able to cope if they test positive for HIV. This acknowledgment of the need for *counseling and support* via the app is consistent with findings of other studies [20,54]. Our findings show that participants are important partners in cocreating a sustainable intervention that will improve the health outcomes of the community.

Study Limitations and Strengths

Owing to the specific study sites, sampling, and sample size, generalizability of the qualitative results to other settings in uMkhanyakude district or outside KZN province may be limited. Some findings presented in this paper were based on a *hypothetical mHealth intervention* described to participants; therefore, assessment of actual receptivity and usage of the app for linkage to care and its impact on HIVST behaviors needs to be conducted after it has been developed, as this may differ from how people imagined it to be. Owing to the ever-changing nature of mobile technology and user expectations in app-based communications, the specific recommendations reported in this paper about the proposed app for HIVST promotion and self-administration may have time-limited relevance. Nevertheless, the preformative and main research data gathered from 2 different communities highlight the strength of the study, as it afforded us the opportunity to draw on a range of experiences across 2 sites. The differences and similarities in the experiences of participants across the sites were invaluable and add nuance to our understanding of the current situation, as well as the potential to codevelop (with the community including health care providers) a suitable mHealth intervention to improve people's health outcomes in a resource-constrained setting.

Discussion

Overall, the study shows that mHealth intervention developed with potential end users, including health care providers, may allow people to explore and test their candidacy for HIV testing and treatment or prevention. However, it cannot overcome the fear of an *HIV identity* (self-stigmatization) and other factors (eg, external stigma) that permeate people's everyday lives, making it difficult for them to assert their candidacy without improving the quality of HIV care available. Features of HIVST and mobile phone–connected Web-based clinical pathways that have the potential to overcome some of the social costs of HIV testing and treatment are their privacy and convenience, as well as allowing the user to be in control of the process. However, there are residual anxieties around digital literacy and the potential to increase health inequalities. mHealth interventions' greatest potential will be to support decentralizing HIV testing, care, and prevention services from clinics to key places that are male- and youth-friendly in the community or workplace where men or young people can be reached. To ensure its effectiveness in reducing HIV incidence and mortality, work still needs to be done to tackle the social norms that continue to fuel stigma and the fear of an *HIV identity*.

Acknowledgments

The authors would like to acknowledge the community of uMkhanyakude district and the participants for their time. Our sincere appreciation goes to Nondumiso Dlamini, Sibongiseni Xulu, Thembani Mhlongo, and Manono Luthuli for participant recruitment, data collection, processes, and overall coordination of the research project. The authors are also grateful to Kobus Herbst, Valerian

Turbe, Michael Thomas, Pam Sonnenberg, Deenan Pillay, and our funders. m-Africa is funded by a Medical Research Council GCRF Global Infections Foundation Award (MR/P024378/1) and is part of the EDCTP2 program, which is supported by the European Union. The research leading to these results has received funding from the People Program (Marie Curie Actions) of the European Union's seventh Framework Program FP7/2007-2013, under REA grant agreement no. 612216. Africa Health Research Institute is supported by core funding from the Wellcome Trust (Core grant number 082384/Z/07/Z).

Authors' Contributions

MS and RM conceived the study. OA, MS, and CH designed the study. OA, MS, and JS wrote the first draft of the manuscript. OA, MS, JS, CH, RM, AB, and CE read and critically revised the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. In Resolution adopted by the General Assembly: United Nations. 2015. United Nations: Transforming our world: The 2030 agenda for sustainable development URL: https://www.un.org/en/development/desa/population/migration/generalassembly/docs/globalcompact/A_RES_70_1_E.pdf
2. Bor J, Herbst AJ, Newell M, Bärnighausen T. Increases in adult life expectancy in rural South Africa: valuing the scale-up of HIV treatment. Science 2013 Feb 22;339(6122):961-965 [FREE Full text] [doi: 10.1126/science.1230413] [Medline: 23430655]
3. Tanser F, Bärnighausen T, Grapsa E, Zaidi J, Newell M. High coverage of ART associated with decline in risk of HIV acquisition in rural KwaZulu-Natal, South Africa. Science 2013 Feb 22;339(6122):966-971 [FREE Full text] [doi: 10.1126/science.1228160] [Medline: 23430656]
4. The Human Sciences Research Council. 2018. The Fifth South African National Hiv Prevalence, Incidence, Behaviour and Communication Survey, 2017 URL: http://www.hsrc.ac.za/uploads/pageContent/9234/SABSSMV_Impact_Assessment_Summary_ZA_ADS_cleared_PDFA4.pdf [accessed 2018-12-10]
5. Iwuji CC, Orne-Gliemann J, Larmarange J, Balestre E, Thiebaut R, Tanser F, et al. The impact of universal test and treat on HIV incidence in a rural South African population: ANRS 12249 TasP trial, 2012-2016. J Int AIDS Soc 2016;19(Suppl 5):104-105 [FREE Full text]
6. Adeagbo O, Herbst C, Dlamini N, Mhlongo T, Blandford A, Valerian T, et al. KwaZulu-Natal Department of Health. 2018. The 'Missing Middle' in HIV Treatment Cascade: Ensuring Men's and Young People's access to HIV Care in KwaZulu-Natal, South Africa URL: http://www.kznhealth.gov.za/research/Research-Day-2018-Abstract-Booklet.pdf [accessed 2018-12-10]
7. Chikovore J, Gillespie N, McGrath N, Orne-Gliemann J, Zuma T, ANRS 12249 TasP Study Group. Men, masculinity, and engagement with treatment as prevention in KwaZulu-Natal, South Africa. AIDS Care 2016;28(Suppl 3):74-82 [FREE Full text] [doi: 10.1080/09540121.2016.1178953] [Medline: 27421054]
8. Cohen M, Chen Y, McCauley M, Gamble T, Hosseinipour M, Kumarasamy N, HPTN 052 Study Team. Antiretroviral therapy for the prevention of HIV-1 transmission. N Engl J Med 2016 Sep 1;375(9):830-839 [FREE Full text] [doi: 10.1056/NEJMoa1600693] [Medline: 27424812]
9. INSIGHT START Study Group, Lundgren JD, Babiker AG, Gordin F, Emery S, Grund B, et al. Initiation of antiretroviral therapy in early asymptomatic HIV infection. N Engl J Med 2015 Aug 27;373(9):795-807 [FREE Full text] [doi: 10.1056/NEJMoa1506816] [Medline: 26192873]
10. TEMPRANO ANRS 12136 Study Group, Danel C, Moh R, Gabillard D, Badje A, Le Carrou J, et al. A trial of early antiretrovirals and isoniazid preventive therapy in Africa. N Engl J Med 2015 Aug 27;373(9):808-822. [doi: 10.1056/NEJMoa1507198] [Medline: 26193126]
11. Iwuji CC, Orne-Gliemann J, Larmarange J, Balestre E, Thiebaut R, Tanser F, ANRS 12249 TasP Study Group. Universal test and treat and the HIV epidemic in rural South Africa: a phase 4, open-label, community cluster randomised trial. Lancet HIV 2018 Mar;5(3):e116-e125. [doi: 10.1016/S2352-3018(17)30205-9] [Medline: 29199100]
12. Adams AK, Zamberia AM. 'I will take ARVs once my body deteriorates': an analysis of Swazi men's perceptions and acceptability of Test and Start. Afr J AIDS Res 2017 Dec;16(4):295-303. [doi: 10.2989/16085906.2017.1362015] [Medline: 29132279]
13. Seeley J, Bond V, Yang B, Floyd S, MacLeod D, Viljoen L, HPTN 071 (PopART) Study Team. Understanding the time needed to Link to Care and Start ART in Seven HPTN 071 (PopART) study communities in Zambia and South Africa. AIDS Behav 2019 Apr;23(4):929-946 [FREE Full text] [doi: 10.1007/s10461-018-2335-7] [Medline: 30415432]
14. Wilson D, Taaffe J, Fraser-Hurt N, Gorgens M. The economics, financing and implementation of HIV treatment as prevention: what will it take to get there? Afr J AIDS Res 2014;13(2):109-119. [doi: 10.2989/16085906.2014.943254] [Medline: 25174628]
15. Choko A, MacPherson P, Webb E, Ball H, Sambakunsi R, Mdolo A, et al. One Year Outcomes Following Community-Based HIV Self-Testing: A Prospective Study in Malawi. In: Proceedings of the 21st Conference on Retroviruses and Opportunistic

Infections. 2014 Presented at: CROI'14; March 3-6, 2014; Boston, Massachusetts URL: http://www.croiconference.org/sessions/one-year-outcomes-following-community-based-hiv-self-testing-prospective-study-malawi
16. Choko AT, MacPherson P, Webb EL, Willey BA, Feasy H, Sambakunsi R, et al. Uptake, accuracy, safety, and linkage into care over two years of promoting annual self-testing for HIV in Blantyre, Malawi: a community-based prospective study. PLoS Med 2015 Sep;12(9):e1001873 [FREE Full text] [doi: 10.1371/journal.pmed.1001873] [Medline: 26348035]
17. MacPherson P, Lalloo DG, Webb EL, Maheswaran H, Choko AT, Makombe SD, et al. Effect of optional home initiation of HIV care following HIV self-testing on antiretroviral therapy initiation among adults in Malawi: a randomized clinical trial. J Am Med Assoc 2014;312(4):372-379 [FREE Full text] [doi: 10.1001/jama.2014.6493] [Medline: 25038356]
18. Bradway M, Pfuhl G, Joakimsen R, Ribu L, Grøttland A, Årsand E. Analysing mHealth usage logs in RCTs: explaining participants' interactions with type 2 diabetes self-management tools. PLoS One 2018;13(8):e0203202 [FREE Full text] [doi: 10.1371/journal.pone.0203202] [Medline: 30161248]
19. Georgette N, Siedner M, Petty C, Zanoni B, Carpenter S, Haberer J. Impact of a clinical program using weekly Short Message Service (SMS) on antiretroviral therapy adherence support in South Africa: a retrospective cohort study. BMC Med Inform Decis Mak 2017 Feb 20;17(1):18 [FREE Full text] [doi: 10.1186/s12911-017-0413-9] [Medline: 28219372]
20. Zhao Y, Zhu X, Pérez AE, Zhang W, Shi A, Zhang Z, et al. mHealth approach to promote Oral HIV self-testing among men who have sex with men in China: a qualitative description. BMC Public Health 2018 Sep 27;18(1):1146 [FREE Full text] [doi: 10.1186/s12889-018-6046-9] [Medline: 30261856]
21. Hirsch-Moverman Y, Daftary A, Yuengling K, Saito S, Ntoane M, Frederix K, et al. Using mHealth for HIV/TB treatment support in Lesotho: enhancing patient-provider communication in the start study. J Acquir Immune Defic Syndr 2017 Jan 1;74(Suppl 1):S37-S43 [FREE Full text] [doi: 10.1097/QAI.0000000000001202] [Medline: 27930610]
22. Wood CS, Thomas MR, Budd J, Mashamba-Thompson TP, Herbst K, Pillay D, et al. Taking connected mobile-health diagnostics of infectious diseases to the field. Nature 2019 Feb;566(7745):467-474. [doi: 10.1038/s41586-019-0956-2] [Medline: 30814711]
23. Free C, Phillips G, Felix L, Galli L, Patel V, Edwards P. The effectiveness of M-health technologies for improving health and health services: a systematic review protocol. BMC Res Notes 2010 Oct 6;3:250 [FREE Full text] [doi: 10.1186/1756-0500-3-250] [Medline: 20925916]
24. Déglise C, Suggs LS, Odermatt P. Short message service (SMS) applications for disease prevention in developing countries. J Med Internet Res 2012 Jan 12;14(1):e3 [FREE Full text] [doi: 10.2196/jmir.1823] [Medline: 22262730]
25. Better Health. 2011. The Ottawa Charter for Health Promotion URL: https://www.betterhealth.vic.gov.au/health/ServicesAndSupport/ottawa-charter-for-health-promotion [accessed 2019-01-30]
26. Pouchster J, Oates R. Pew Research Center. 2015. Cell Phones in Africa: Communication Lifeline URL: https://www.pewresearch.org/global/wp-content/uploads/sites/2/2015/04/Pew-Research-Center-Africa-Cell-Phone-Report-FINAL-April-15-2015.pdf [accessed 2018-09-30]
27. Health Systems Trust. 2015. mHealth Strategy 2015-2019 URL: http://www.hst.org.za/publications/NonHST%20Publications/mHealth%20Strategy%202015.pdf [accessed 2018-12-30]
28. Ospina-Pinillos L, Davenport T, Diaz AM, Navarro-Mancilla A, Scott E, Hickie IB. Using participatory design methodologies to co-design and culturally adapt the Spanish version of the mental health eClinic: qualitative study. J Med Internet Res 2019 Aug 2;21(8):e14127 [FREE Full text] [doi: 10.2196/14127] [Medline: 31376271]
29. Dixon-Woods M, Cavers D, Agarwal S, Annandale E, Arthur A, Harvey J, et al. Conducting a critical interpretive synthesis of the literature on access to healthcare by vulnerable groups. BMC Med Res Methodol 2006 Jul 26;6:35 [FREE Full text] [doi: 10.1186/1471-2288-6-35] [Medline: 16872487]
30. Macdonald S, Blane D, Browne S, Conway E, Macleod U, May C, et al. Illness identity as an important component of candidacy: contrasting experiences of help-seeking and access to care in cancer and heart disease. Soc Sci Med 2016 Nov;168:101-110. [doi: 10.1016/j.socscimed.2016.08.022] [Medline: 27643844]
31. Mackenzie M, Conway E, Hastings A, Munro M, O'Donnell C. Is 'candidacy' a useful concept for understanding journeys through public services? A critical interpretive literature synthesis. Soc Policy Adm 2013;47(7):806-825. [doi: 10.1111/j.1467-9515.2012.00864.x]
32. Nkosi B, Seeley J, Ngwenya N, Mchunu SL, Gumede D, Ferguson J, et al. Exploring adolescents and young people's candidacy for utilising health services in a rural district, South Africa. BMC Health Serv Res 2019 Mar 28;19(1):195 [FREE Full text] [doi: 10.1186/s12913-019-3960-1] [Medline: 30922372]
33. Klassen A, Smith K, Shariff-Marco S, Juon HS. A healthy mistrust: how worldview relates to attitudes about breast cancer screening in a cross-sectional survey of low-income women. Int J Equity Health 2008 Jan 31;7:5 [FREE Full text] [doi: 10.1186/1475-9276-7-5] [Medline: 18237395]
34. Stern E, Rau A, Cooper D. Sexual and reproductive health perceptions and practices as revealed in the sexual history narratives of South African men living in a time of HIV/AIDS. SAHARA J 2014;11:233-244 [FREE Full text] [doi: 10.1080/17290376.2014.985701] [Medline: 25495581]
35. Skovdal M, Campbell C, Nyamukapa C, Gregson S. When masculinity interferes with women's treatment of HIV infection: a qualitative study about adherence to antiretroviral therapy in Zimbabwe. J Int AIDS Soc 2011 Jun 9;14:29 [FREE Full text] [doi: 10.1186/1758-2652-14-29] [Medline: 21658260]

36. Siu GE, Wight D, Seeley JA. Masculinity, social context and HIV testing: an ethnographic study of men in Busia district, rural eastern Uganda. BMC Public Health 2014 Jan 13;14:33 [FREE Full text] [doi: 10.1186/1471-2458-14-33] [Medline: 24417763]
37. DiCarlo AL, Mantell JE, Remien RH, Zerbe A, Morris D, Pitt B, et al. 'Men usually say that HIV testing is for women': gender dynamics and perceptions of HIV testing in Lesotho. Cult Health Sex 2014;16(8):867-882 [FREE Full text] [doi: 10.1080/13691058.2014.913812] [Medline: 24854495]
38. van Heerden A, Msweli S, van Rooyen H. 'Men don't want things to be seen or known about them': a mixed-methods study to locate men in a home based counselling and testing programme in KwaZulu-Natal, South Africa. Afr J AIDS Res 2015 Dec 20;14(4):353-359. [doi: 10.2989/16085906.2015.1121881]
39. Zissette S, Watt MH, Prose NS, Mntambo N, Moshabela M. 'If you don't take a stand for your life, who will help you?': men's engagement in HIV care in KwaZulu-Natal, South Africa. Psychol Men Masc 2016 Jul;17(3):265-273 [FREE Full text] [doi: 10.1037/men0000025] [Medline: 27616937]
40. Tabana H, Doherty T, Rubenson B, Jackson D, Ekström AM, Thorson A. 'testing together challenges the relationship': consequences of HIV testing as a couple in a high HIV prevalence setting in rural South Africa. PLoS One 2013;8(6):e66390 [FREE Full text] [doi: 10.1371/journal.pone.0066390] [Medline: 23824067]
41. UNAIDS. 2017. Reaching Out to Men and Boys: Addressing a Blind Spot in the Response to HIV URL: http://www.unaids.org/sites/default/files/media_asset/blind_spot_en.pdf [accessed 2018-08-14]
42. Dovel K, Yeatman S, Watkins S, Poulin M. Men's heightened risk of AIDS-related death: the legacy of gendered HIV testing and treatment strategies. AIDS 2015 Jun 19;29(10):1123-1125 [FREE Full text] [doi: 10.1097/QAD.0000000000000655] [Medline: 26035315]
43. Dovel K, Yeatman S, Watkins S, Poulin M. Prioritizing strategies to reduce AIDS-related mortality for men in sub-Saharan Africa: authors' reply. AIDS 2016 Jan 2;30(1):158-159. [doi: 10.1097/QAD.0000000000000898] [Medline: 26731761]
44. Pai NP, Behlim T, Abrahams L, Vadnais C, Shivkumar S, Pillay S, et al. Will an unsupervised self-testing strategy for HIV work in health care workers of South Africa? A cross sectional pilot feasibility study. PLoS One 2013;8(11):e79772 [FREE Full text] [doi: 10.1371/journal.pone.0079772] [Medline: 24312185]
45. Huerga H, van Cutsem G, Farhat J, Puren A, Bouhenia M, Wiesner L, et al. Progress towards the UNAIDS 90-90-90 goals by age and gender in a rural area of KwaZulu-Natal, South Africa: a household-based community cross-sectional survey. BMC Public Health 2018 Mar 2;18(1):303 [FREE Full text] [doi: 10.1186/s12889-018-5208-0] [Medline: 29499668]
46. Carrasco MA, Fleming P, Wagman J, Wong V. Toward 90-90-90: identifying those who have never been tested for HIV and differences by sex in Lesotho. AIDS Care 2018 Mar;30(3):284-288. [doi: 10.1080/09540121.2017.1372559] [Medline: 28868903]
47. Wirtz AL, Clouse E, Veronese V, Thu KH, Naing S, Baral SD, et al. New HIV testing technologies in the context of a concentrated epidemic and evolving HIV prevention: qualitative research on HIV self-testing among men who have sex with men and transgender women in Yangon, Myanmar. J Int AIDS Soc 2017 Apr 25;20(1):21796 [FREE Full text] [doi: 10.7448/IAS.20.01.21796] [Medline: 28453242]
48. Frye V, Wilton L, Hirshfied S, Chiasson MA, Usher D, Lucy D, All About Me Study Team. 'Just because it's out there, people aren't going to use it.' HIV self-testing among young, black MSM, and transgender women. AIDS Patient Care STDS 2015 Nov;29(11):617-624 [FREE Full text] [doi: 10.1089/apc.2015.0100] [Medline: 26376029]
49. Wood BR, Ballenger C, Stekler JD. Arguments for and against HIV self-testing. HIV AIDS (Auckl) 2014;6:117-126 [FREE Full text] [doi: 10.2147/HIV.S49083] [Medline: 25114592]
50. Scambler G. Health-related stigma. Sociol Health Illn 2009 Apr;31(3):441-455 [FREE Full text] [doi: 10.1111/j.1467-9566.2009.01161.x] [Medline: 19366430]
51. Krause J, Subklew-Sehume F, Kenyon C, Colebunders R. Acceptability of HIV self-testing: a systematic literature review. BMC Public Health 2013 Aug 8;13:735 [FREE Full text] [doi: 10.1186/1471-2458-13-735] [Medline: 23924387]
52. Kirby T, Thornber-Dunwell M. Phone apps could help promote sexual health in MSM. Lancet 2014 Oct 18;384(9952):1415. [doi: 10.1016/S0140-6736(14)61849-3] [Medline: 25390311]
53. Young SD, Cumberland WG, Nianogo R, Menacho LA, Galea JT, Coates T. The HOPE social media intervention for global HIV prevention in Peru: a cluster randomised controlled trial. Lancet HIV 2015 Jan;2(1):e27-e32 [FREE Full text] [doi: 10.1016/S2352-3018(14)00006-X] [Medline: 26236767]
54. Zhang X, Wen D, Liang J, Lei J. How the public uses social media WeChat to obtain health information in China: a survey study. BMC Med Inform Decis Mak 2017 Jul 5;17(Suppl 2):66 [FREE Full text] [doi: 10.1186/s12911-017-0470-0] [Medline: 28699549]
55. Li X, Xu Z, Tang N, Ye C, Zhu X, Zhou T, et al. Effect of intervention using a messaging app on compliance and duration of treatment in orthodontic patients. Clin Oral Investig 2016 Nov;20(8):1849-1859. [doi: 10.1007/s00784-015-1662-6] [Medline: 26631059]

Abbreviations

ART: antiretroviral therapy

CA: first community
CB: second community
FGD: focus group discussion
HIVST: HIV self-testing
IDI: in-depth interview
KZN: KwaZulu-Natal
mHealth: mobile health
MSM: men who have sex with men
PrEP: Pre-Exposure Prophylaxis
SA: South Africa
SP: service provider
TasP: treatment as prevention
UTT: Universal HIV test and treat

Edited by G Eysenbach; submitted 29.07.19; peer-reviewed by C Agboola, J Marcano-Belisario; comments to author 31.08.19; revised version received 06.09.19; accepted 07.09.19; published 18.11.19.

Please cite as:
Adeagbo O, Herbst C, Blandford A, McKendry R, Estcourt C, Seeley J, Shahmanesh M
Exploring People's Candidacy for Mobile Health–Supported HIV Testing and Care Services in Rural KwaZulu-Natal, South Africa: Qualitative Study
J Med Internet Res 2019;21(11):e15681
URL: https://www.jmir.org/2019/11/e15681
doi:10.2196/15681
PMID:31738174

Original Paper

Using Social Media to Uncover Treatment Experiences and Decisions in Patients With Acute Myeloid Leukemia or Myelodysplastic Syndrome Who Are Ineligible for Intensive Chemotherapy: Patient-Centric Qualitative Data Analysis

Alison Booth[1], MSc; Timothy Bell[2], MHA; Sonia Halhol[1], MSc; Shiyu Pan[1], MSc; Verna Welch[2], PhD, MPH; Evie Merinopoulou[1], MSc; Dimitra Lambrelli[1], PhD; Andrew Cox[1], PhD

[1]Evidera, London, United Kingdom
[2]Pfizer, New York, NY, United States

Corresponding Author:
Alison Booth, MSc
Evidera
The Ark, 2nd Floor
201 Talgarth Road
London, W6 8BJ
United Kingdom
Phone: 44 (0)208 576 5048
Email: alison.booth@evidera.com

Abstract

Background: Until recently, treatment options were limited for patients with acute myeloid leukemia and myelodysplastic syndrome (AML and MDS) who are ineligible for intensive chemotherapy. Owing to the condition's rapid progression, it is difficult to identify what is most important to patients when making treatment decisions. Patients' needs can be better addressed by gaining a deeper understanding of their perspectives, which is valuable in the decision-making process. The Food and Drug Administration recently encouraged the use of social media as a tool to gain insight on patients' perspectives regarding symptoms experienced and the impacts of their disease.

Objective: This study aimed to use disease-specific social media posts by patients with AML or MDS who are ineligible for intensive chemotherapy and their caregivers to capture factors they feel are most important, and to provide current evidence to inform and characterize these perspectives.

Methods: Posts by patients with AML or MDS and their caregivers were extracted from publicly available discussions on 3 large AML- or MDS–specific sites. These posts were manually reviewed to only include patients who are ineligible for intensive chemotherapy. A total of 1443 posts from 220 AML patients/caregivers and 2733 posts from 127 MDS patients/caregivers met the study inclusion criteria. A qualitative data analysis (QDA) of a sample of 85 patients'/caregivers' posts was conducted to identify themes, and a targeted QDA of posts from 79 users focused on treatment decision discussions. Posts were manually reviewed, and relevant text segments were coded and grouped into categories and overall themes.

Results: Eighty-six percent (73/85) of users in the overall QDA had relevant information about the key objectives. The most commonly discussed treatment experience theme was the humanistic burden of AML or MDS in terms of emotional/physical impact and impact on family (86%, 63/73 of users), followed by treatment decisions (56%, 41/73) and unmet needs (50%, 37/73). In the QDA of treatment decisions, 60 posts from 45 users contained relevant information. Patients commonly reported the desire to reach specific milestones, including birthdays and weddings. They wished for a better quality of life over quantity of life, did not want the risk of suffering from side effects, and expressed a clear preference to be at home rather than in a hospital or care home.

Conclusions: This study was a novel application of disease-specific social media. It highlighted experiences in the current treatment of AML and MDS, including information gaps, patient/caregiver uncertainty, and the importance of understanding patients'/caregivers' goals and opinions. A clear finding from this research was the importance of reaching certain personal life goals and being at home with family and friends. The analysis showed that patients/caregivers face additional challenges, including humanistic impacts and a lack of information regarding treatment options.

(J Med Internet Res 2019;21(11):e14285) doi:10.2196/14285

KEYWORDS

social media; health-related quality of life; patient-centric; leukemia; myeloid; acute; myelodysplastic syndromes; natural language processing; patient preference; qualitative research

Introduction

Background

Myelodysplastic syndromes (MDSs) constitute a heterogenous group of hematopoietic stem cell disorders [1] in which the bone marrow produces immature white blood cells, red blood cells, platelets, or a combination of all three. MDS can be rapidly progressive and is associated with a risk of evolution into acute myeloid leukemia (AML) [2]. AML, an orphan disease, is an aggressive cancer of the blood, consisting of a group of relatively well-defined hematopoietic neoplasms [1]. The disease begins as abnormal blood cells that are produced in the bone marrow and can spread throughout the circulatory system and beyond if not diagnosed and treated quickly [3]. Both AML and MDS are hematologic diseases that generally affect older adults; they are uncommon before the ages of 45 and 50 years and are most commonly diagnosed in patients who are in their late 60s or 70s [4,5].

The annual incidence of AML in the United States is approximately 3.75 cases per 100,000 people, and the incidence rate of MDS is approximately 3.30 cases per 100,000 persons [1]. White men have the highest incidence of MDS (6.9 cases per 100,000 people) [6]. In 2018, there were an estimated 19,520 new cases of AML, and there are approximately 14,275 new cases of MDS annually according to data from 2010 to 2014 in the United States [6]. The number of new cases diagnosed each year is likely to increase as the size of the elderly population increases.

Prognosis is poor for patients with AML and MDS, and effective treatment options are limited. Prognosis for AML generally decreases rapidly with age—5-year survival in adults is estimated to be around 24%; however, this drops to less than 5% in individuals aged 65 years and older [7]. The treatment approach for MDS depends on several factors, including the type and risk group of MDS and the patients' age and overall health. Most diagnosed patients are older or in poor health and, thus, are not suitable candidates for a stem cell transplant, which is potentially the only cure for MDS. Alternative treatments for MDS, such as supportive care (transfusions or blood cell growth factors) along with chemotherapy, serve only to relieve symptoms and avoid complications and side effects. The standard treatment for patients with AML is intensive chemotherapy, typically involving cytarabine-based regimens, with the goal being to achieve complete remission. Intensive chemotherapy commonly requires prolonged hospitalization as it is associated with severe myelosuppression and an increased risk of treatment-related mortality. As a consequence, elderly patients with AML who present with poor performance status and/or comorbidities are generally not considered candidates for intensive chemotherapy because the risks are thought to outweigh the benefits. Until recently, treatment options for these patients were limited to clinical trials, nonintensive agents (hypomethylating agents or low-dose cytarabine), or best supportive care [8]. Reported treatment patterns in older AML patients vary; however, estimates using data from 2000 to 2014 suggest that only 40% to 60% of these patients receive any therapy for their disease [9], in part, because of a lack of options, but other reasons are not well understood [10,11].

Owing to poor prognosis and limited treatment options for patients with AML and MDS ineligible for intensive chemotherapy, there are corresponding unmet medical needs. These conditions also impair patients' health-related quality of life (HRQoL) and are associated with high symptom burden [12]. Patients' needs can be better addressed by gaining a deeper understanding of their perspectives on the disease, treatment options, and the impact of both on their lives and those of their caregivers. This information is valuable in the decision-making process—including end-of-life treatment decisions—but is challenging to capture using traditional methods or data sources (eg, electronic medical records and patient questionnaires) because of the rapid progression of the disease and patients' age at diagnosis. Recent studies have highlighted the need to evaluate patient-reported outcomes for symptoms of AML or MDS to outline areas of focus for clinicians and researchers [13-15]. However, studies assessing experiences, preferences, and perspectives regarding treatment decisions for patients with AML or MDS are scarce, especially for those who are ineligible for intensive chemotherapy [16,17]. Those studies are crucial to better inform drug discovery and regulatory decisions and imperative for improving the patient experience.

Disease-specific platforms (eg, patient forums and discussion boards) on social media represent a virtual community where patients and their caregivers spontaneously share experiences and perspectives related to their disease and treatments. This information can provide an instant snapshot of the current humanistic and economic burden of illness in patients with AML or MDS and provide insight into the rationale for patients' treatment decisions. The Food and Drug Administration reinforced this idea in June 2018, when it made recommendations encouraging the use of social media to shed light on the patient's perspectives regarding symptoms and impacts of their diseases, stressing the opportunity to inform medical product development and enhance regulatory decision making [18].

Objectives

The objective of this study was to use social media on disease-specific platforms to identify themes related to disease and treatment experiences of patients with AML or MDS who are ineligible for intensive chemotherapy and their caregivers. It aimed to capture the most important factors that drive patients' treatment decisions by identifying patients' priorities and their reasons for pursuing certain types of treatments over others, including preferences in end-of-life situations.

Methods

Data Source

An initial search of disease-specific social media was conducted, focusing on the feasibility of addressing the study objectives (finalized on April 23, 2018). The search strategy focused on identifying AML and MDS–specific social media forums and discussion boards, using key search terms such as *AML patient forum* and *MDS patient discussion*. An exhaustive list of search terms used can be found in Multimedia Appendix 1.

Generic social media sites such as Facebook, Google+, and Twitter were not considered because of the added complication of filtering out irrelevant material and because access to data is governed by complex terms and conditions and data access fees. Searches were conducted using US and UK Google search engines, the first 5 pages of results were screened by title, and relevant forums were summarized. The searches yielded 41 sites for consideration for use in the study, which were assessed based on the following criteria:

- Each user has a username within the site, and the user can be easily classified as either a patient with AML or MDS or their caregiver/family member. The site topic must be clear, and posts must be specific to AML or MDS.
- The information contained in the forum must be relevant to the key study objectives (ie, disease symptoms and impacts on patients' HRQoL, treatment sequence, treatment patterns, adverse events [AEs], and impact on caregivers).
- The forum site must be active and contain posts from recent years.
- The information must be posted in English.
- Material is freely available for anyone to find and read, and no registration is required.
- Posted content must be programmatically retrievable, without any restrictions from the site's data protection terms and conditions.

Of the 41 sites identified during the initial search, 19 met the selection criteria. The following 3 forum sites with the highest number of posts were selected for this study:

1. AML Online Support Group: an active discussion group for patients with AML—visitors to the site are mainly from India, the United States, Pakistan, Indonesia, and Canada [19,20].
2. MacMillan Cancer Support: AML Online Forum: a support and discussion group for AML patients—visitors to the site are predominantly located in the United Kingdom but are also in the United States, India, Canada, and Australia [21].
3. MDS Foundation: MDS Patient Message Board: a platform where patients are encouraged to ask questions and get answers from fellow patients or MDS experts—the majority of its visitors are from the United States [22].

Social Media Content Extraction and Data Preparation

Data extraction algorithms specific to each site were written in the R computational programming language (version 3.5.1 [23]) to extract posts programmatically from each site. Data extracted included username, post content, thread title, date of the post, and URL. A total of 44,456 posts (AML=21,654 and MDS=22,802) were extracted from 7811 users (AML=6373 and MDS=1438).

All data were programmatically deidentified at the time of extraction, and postal codes and addresses, email addresses, Web links, and telephone numbers were all removed. Following extraction, the data were processed and cleaned. Duplicate posts and non-Unicode Transformation Format-8 characters were removed. Misspellings were corrected using a dictionary of known misspellings in disease-related social media data compiled by the authors. The R hunspell package [24] was then used to identify additional misspellings. Despite these efforts, it is possible that some misspellings remained in the final data. Data were then assembled into individual user posting history files in .txt file format. These files contained all post content from each user chronologically. Extracted posts with the username *deleted_user* could not be assigned to user posting histories. Therefore, these were retained as individual posts and assumed to be individual users.

Study Population

This study focused on patients with AML or MDS who were ineligible for intensive chemotherapy and their caregivers. The inclusion and exclusion criteria are provided in Textboxes 1 and 2, respectively. Intensive and nonintensive chemotherapies are listed in Textbox 3.

Textbox 1. Inclusion criteria.

- Patients/caregivers of patients with acute myeloid leukemia/myelodysplastic syndrome who mention that they receive a nonintensive chemotherapy
- Patients/caregivers of patients with acute myeloid leukemia/myelodysplastic syndrome who mention that they are not treated with intensive chemotherapy
- Patients/caregivers of patients with acute myeloid leukemia/myelodysplastic syndrome who mention that they are treated with an intensive chemotherapy at a low dose
- Patients/caregivers of patients with acute myeloid leukemia/myelodysplastic syndrome who mention that they are ineligible for intensive chemotherapy

Textbox 2. Exclusion criteria.

- Patients/caregivers of patients with acute myeloid leukemia/myelodysplastic syndrome who mention that they are treated with an intensive chemotherapy and who do not mention dose
- Patients/caregivers of patients with acute myeloid leukemia/myelodysplastic syndrome who mention that they have been treated with an intensive chemotherapy and who do not mention that they are no longer eligible to receive it
- Patients/caregivers of patients with acute myeloid leukemia/myelodysplastic syndrome for whom treatment received is not clear

Textbox 3. Chemotherapy treatments used to treat acute myeloid leukemia and myelodysplastic syndrome by intensity used in consideration of inclusion/exclusion criteria. Brand names are listed in parentheses for each generic drug.

Intensive chemotherapy treatments

- Amsacrine (Amsidine)
- Arsenic trioxide (Trisenox)
- Cytarabine (cytosine arabinoside and high-dose cytarabine)
- Daunorubicin (Cerubidin)
- Etoposide (Eposin, Etopophos, and VePesid)
- Fludarabine (Fludara)
- Gemtuzumab ozogamicin (Mylotarg) + intensive chemotherapy
- Idarubicin (Zavedos)
- Mitoxantrone
- Tioguanine (thioguanine and 6-tioguanine)
- Tretinoin (Vesanoid and all-trans retinoic acid)

Nonintensive chemotherapy treatments

- Azacitidine (Vidaza)
- Decitabine (Dacogen)
- Gemtuzumab ozogamicin (Mylotarg)
- Low-dose cytarabine

Sample Size and Saturation of Emerging Themes

Saturation (the point at which no new significant themes emerge from the analysis) was judged and assessed by the research team. At various points throughout the analysis, the team discussed and decided to either add more samples (patient posting histories) from the generated pool of 347 study-eligible patients or stop those analyses if it was felt that saturation had been reached. This was independently done on an objective-by-objective basis.

User Classification

From the user posting histories compiled, an initial sample of posts from approximately 11.42% (892/7811) of randomly selected users was manually reviewed to determine whether the user met the inclusion criteria. The decision to include or exclude a user was based on phrases in post content (ie, “I have AML and am not eligible for intensive chemo”), and discussion of receiving one of the therapies is listed in Textbox 3.

User Selection

For the identification of themes related to disease and treatment experiences of patients, the research team judged that saturation had been reached and stopped further analysis at 80 AML user posting histories and 5 MDS user posting histories. Search terms relevant to end-of-life decisions were generated (details can be found in Table 1) to capture the most important factors that drive treatment decisions for patients, focusing on those that would help address the study objectives; posts containing any 1 of these search terms were included. For the research objective on treatment decisions, it was felt that saturation had not been reached. Therefore, posts of users included following the manual classification were further searched to maximize the relevant content available for analysis, and only those that contained specific search terms listed in Table 1 were included (79 users).

Table 1. Search terms used in targeted qualitative analysis.

Category	Search terms
End-of-life	*end of life, palliative, hospice, comfort care, and passed away*
Treatment	*no more treatment, stop treatment, not having any treatment, stopped treatment, no longer having treatment, and withdrawn*
Chemotherapy	*no more chemo, stop chemo, not having any chemo, stopped chemo, and no longer having chemo*
Living longer	*quality of life, keep living, live longer, stop living, determined, worth living, keep fighting, living for, more options, affairs sorted, come home, came home, be at home, and now at home*

Analysis

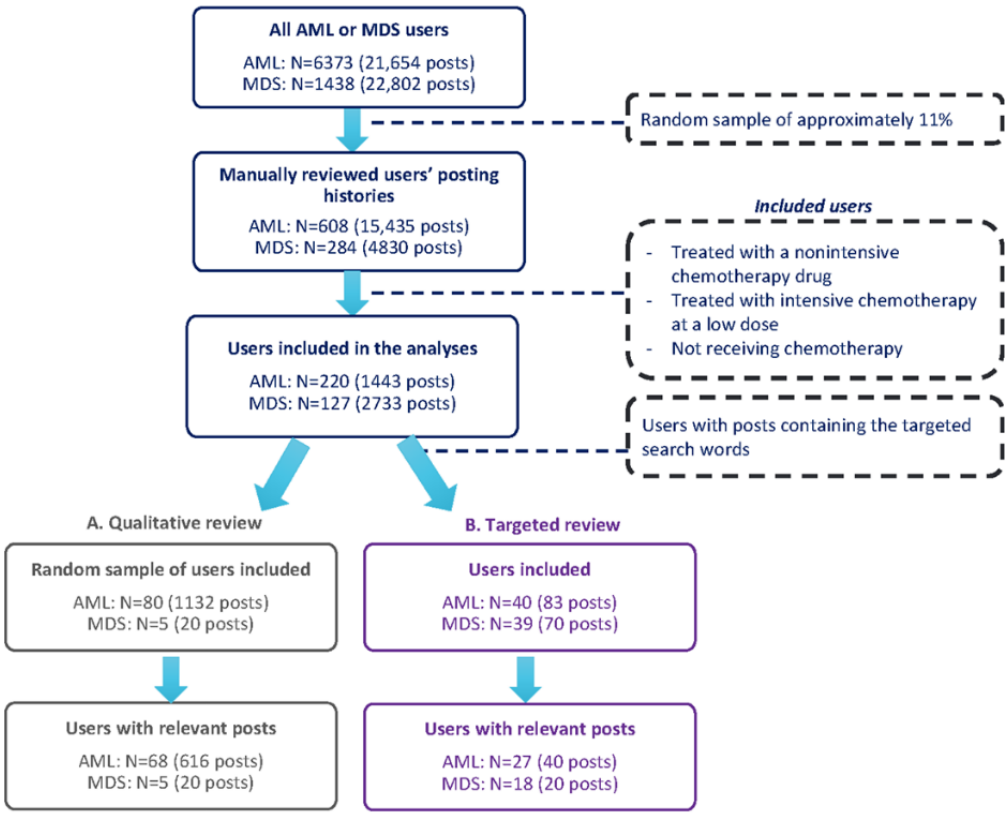
Following the manual review of poster history files, qualitative data analysis (QDA) was undertaken for each of the objectives. Posts were qualitatively reviewed in the R Qualitative Data Analysis Package. Thematic analysis was used, taking a top-down approach to data, using hypothesis coding (themes and codes determined a priori) to develop a coding dictionary based on the key objectives [25]. Text segments relevant to the objectives were highlighted, and open coding was used to add codes based on reviewed data. The occurrence of themes was summarized using descriptive statistics.

Results

Study Population

The posting histories of 892 users (AML=608 and MDS=284) including 20,265 posts (AML=15,435 and MDS=4830) were manually reviewed against the inclusion and exclusion criteria of the study, and, ultimately, 347 users (AML=220 and MDS=127) and 4176 posts (AML=1443 and MDS=2733) were included in the study. For the qualitative review, the 85 users included in the study contributed 1152 posts (AML=1132 and MDS=20). Of the 85 users, 73 (AML=68 and MDS=5) had posts relevant to the study. For the targeted qualitative analysis, 79 users (AML=40 and MDS=39) had posts containing the search terms, and on manual review of posts, 45 users (AML=27 and MDS=18) had posts relevant to the study objectives. For further details, see Figure 1.

Figure 1. Study flow. AML: acute myeloid leukemia; MDS: myelodysplastic syndrome.



Identified Themes Related to Disease and Treatment Experiences

The following 5 overarching themes were identified from the results of the QDA: humanistic burden, treatment decisions, unmet needs, life milestones, and economic burden. The number and proportion of users mentioning each theme are shown in Table 2.

Table 2. Themes mentioned by the largest proportion of users.

High-level themes/subthemes	Description	Users discussing themes, n (%)
Overall QDA[a] (N=73)		
Humanistic burden	Impact on social life, impact on family, inability to do day-to-day activities, physical impact, and psychological impact	63 (86)
Treatment decision[b]	Reasons for declining or not having treatment, reasons for pursuing one treatment over another, and stakeholders involved in the treatment decision	41 (56)
Unmet needs	Emotional support, management of symptoms and side effects, and treatment options	37 (50)
Life milestones	Spending time at home during the end-of-life period, putting affairs in order, and family and social events	14 (19)
Economic burden	Medical expenses, travel expenses, and inability to work	9 (12)
Treatment decision–targeted QDA (N=45)		
Health-related quality of life	Severity of side effects, ability to leave hospital, ability to travel, and being able to socialize	20 (44)
Home and family	Being at home, reaching specific events, and spending time with family	19 (42)
Physician decision	What the physician thinks is the best treatment option for the patient	5 (11)
Patient and family wishes	What the patient and their family want to do regarding treatment	4 (9)
Affairs in order	Finances, childcare, and writing wills	4 (9)

[a]QDA: qualitative data analysis.

[b]A total of 16 users included in the overall QDA had posts in the targeted treatment decision search; therefore, the total number of users mentioning the treatment decision theme in the study was 70. However, as no targeted search was conducted for the other themes, this number is not compared in the table above.

Humanistic Burden

The humanistic burden of AML and MDS was felt strongly by patients and caregivers and sometimes affected daily activities and HRQoL. Patients with AML or MDS were mostly burdened by physical impacts of the disease and/or treatment and most frequently mentioned feeling fatigue or weakness, followed by experiencing infections and fever. Patients experienced AEs that included coughing, headaches, loss of appetite (and subsequent weight loss), nausea and vomiting, pain, mouth blisters, nose bleeds, and hair loss. Many of these impacted daily activities such as eating, sleeping, and going out:

> *My mother is very frail and suffers with fatigue. She also has an infection so her doctors are holding off on any chemotherapy.* [Caregiver]

Posts by patients with AML or MDS revealed various psychological and emotional impacts of the disease. On diagnosis, many patients felt scared and anxious about their upcoming treatment. Many patients and their caregivers were unfamiliar with the disease before diagnosis, and they were left feeling confused because of their lack of knowledge about the disease and treatment options; in turn, this made them more likely to feel desperate and hopeless. The disease and/or treatment sometimes took a toll on patients' cognitive abilities, affecting their concentration and memory as well as their mood:

> *I am very anxious at the moment and am not sleeping because I am waiting for my test results. I'm scared that I may not have much longer left.* [Patient]

> *Every day I feel fatigued, cannot concentrate and generally feel down. I've also begun to struggle with anger.* [Patient]

> *When my treatment stopped I was on the verge of having a breakdown, and since then have also been diagnosed with depression and PTSD. I also struggle sleeping and with nightmares so I am taking anti-depressants.* [Patient]

Family members felt an increasing responsibility to share caring duties and to provide emotional and financial support for patients with AML or MDS. The emotional impact on family was described as heartache, feelings of stress, confusion, fear, and guilt. As caregivers, patients' families sometimes had to assume additional responsibilities such as making decisions on behalf of the patient, looking after the home, handling paperwork, and taking care of the patient and other relatives. This additional pressure on caregivers sometimes impacted their emotional and psychological well-being as well:

> *I have felt desperate and helpless, and I find it difficult to talk to friends about. It's difficult to manage visits to the hospital as well at the same time as working.* [Caregiver]
>
> *I feel that I am useless to my husband, who now has to care for the children, me and manage our finances. I worry that he will resent me.* [Patient]

Treatment Decisions

The results of the QDA further categorized treatment decisions into the following 5 subthemes: HRQoL, home and family, physician decision, patient and family wishes, and putting affairs in order. The number and proportion of users mentioning each theme are provided in Table 2. The themes mentioned by the largest proportion of users were HRQoL and home and family (20/45, 44% and 19/45, 42%, respectively). The themes mentioned less commonly were patient and family wishes and putting affairs in order (both were 4/45, 9%).

Patients who discussed reasons for declining or not receiving chemotherapy and the stakeholders involved in the decision commonly mentioned avoiding potential AEs to maintain HRQoL as an important factor in their decisions. Patients (and their caregivers) expressed that they would rather have a better HRQoL over quantity of life and did not want to take the risk of suffering from side effects in the time they had left; as a result, some patients preferred not to receive treatment:

> *The doctors decided not to give him more chemotherapy the side-effects outweighed the benefits and threatened his life [...] the physician offered a chemotherapy with no side effects, but he does not want the risk. He is stronger and on good days we walk and go for drives, but he does get tired.* [Caregiver]
>
> *My mom said that the quality of her life is a priority over quantity. because of this she does not want to try the azacitidine[...] it makes me sad but her decision makes sense and I respect it.* [Caregiver]

In many cases, it was the physician who decided that a patient would not receive any treatment. Sometimes this aligned with the wishes of the patient and family; however, in some posts, it was clear that the patient/caregiver disagreed with the physician. They felt that the physicians did not want to fight the disease for the same amount of time as they did:

> *It felt like the doctors gave up before the patient did, so I kept fighting with my father and for my father.* [Caregiver]

For patients who discussed the reasons for pursuing certain types of treatments over others and the stakeholders involved in this decision, the most commonly reported reason for choosing a certain treatment was the desire for the patient to be at home rather than in a hospital or care home; this was expressed by both patients and caregivers. Treatments that allow patients to be at home and offer a comparable prognosis were the preferred option:

> *She has the option to go to a hospital closer to home and be able to go home every day. She said that all she wants is to be able to go home. This chemotherapy treatment will let her go home and give her as good prognosis.* [Caregiver]

As would be expected, physicians' opinions played an important role in pursuing a particular treatment. Caregivers did not always agree with the treatment decision recommended by the physicians, and in some cases, felt that physicians were going too far and compromising the patients' quality of life (QoL). The patient's own wishes regarding their treatment—to explore all options and fight until the very end or to stop treatment—were also an important factor:

> *He had multiple rounds of azacitidine he isn't improving [...] it seems like my father is their experiment, the doctors were offended when I said this. But his health and life quality has gone downhill and I am tired and annoyed and don't know what to do.* [Caregiver]
>
> *My grandmother often felt weak, or was in pain, some days she cried or was confused, but, despite this, she did not succumb to it. Now, she's on a ventilator in the hospital and I am fighting for her wish.* [Caregiver]

Unmet Needs

Patients and their caregivers expressed that they had run out of options or that there were not enough treatments available to them. The lack of treatments for elderly patients was also a topic of discussion, and some patients and caregivers expressed a view that age should not interfere with a patient's ability to obtain treatment:

> *There aren't a lot of solutions and no miracle drug.* [Patient]

Other posts identified the view that there was a lack of information and support from physicians. Some patients and caregivers expressed that they felt like their physicians did not provide adequate details needed for them to make informed decisions regarding treatments. Patients and caregivers also indicated a lack of empathy from physicians, which was perceived by them as the physicians refusing to continue treating the patients:

> *His results are being sent for a second opinion and so we can find a new doctor. our current oncologist isn't helpful or supportive.* [Caregiver]
>
> *The doctor point-blank told her they wouldn't resuscitate if something happened, almost like saying there was no point.* [Caregiver]

Patients and caregivers reported receiving inadequate information about specific treatments. Some mistakenly thought that the treatment would be a potential cure, and this sometimes led to disappointment and uncertainty regarding the effectiveness of the treatment and its impact on survival:

> *I was wrong thinking this was what would cure me, which was a stupid thought, rather than an opportunity to live.* [Patient]
>
> *No one is really sure how this treatment works, everything occurred so fast and wasn't explained*

well, don't know whether it's because she doesn't know how it works. [Caregiver]

Overall, patients and caregivers expressed that there was a lack of information being provided to them regarding prognosis. They were often uncertain about what they had been told and needed clarity on the meaning of *poor prognosis*. Caregivers particularly mentioned a lack of warning about the potential for the disease to progress extremely rapidly and the way this could affect the patient. Some caregivers indicated that they would have made different decisions had they known this:

Doctors said he isn't a candidate for high-dose chemotherapy so instead he's getting a milder dose and he has a poor prognosis. what does this mean? [Caregiver]

I'm trying to research disease progression because we've all been given broad ranges of days, weeks, or months. [Caregiver]

When I went to see her at the hospital she didn't recognize me. I didn't expect this, nor did I think it would happen. [Caregiver]

Life Milestones

Patient priorities and preferences in end-of-life situations revolved around being able to spend time with family/friends and was extremely important to both patients and caregivers. This included making the most of the time left or fulfilling the desire to spend their final moments with loved ones. Many patients wished to live long enough to reach a specific event (eg, wedding, Christmas, and birthday), and less frequently, they wanted to survive a specified amount of time that was not tied to a specific occasion:

It was his wish to live until Christmas and our wedding anniversary. he died in peace at home just when he wanted. [Caregiver]

We have a vacation booked soon and we clearly want to go, it's in another country. it's the only opportunity we get to see our daughter a year. [Patient]

Today was sad because there are no plans to bring my father home. i had hopes and plans of being together as a family for the weekend and things being normal. [Caregiver]

Patients had to balance accomplishing certain tasks with the impact a treatment would have on their mental and physical capacity to conduct such tasks. Many patients prioritized putting their affairs in order in preparation for when they passed away, including finances (selling houses and settling bills), funeral plans, and arrangement for their family members:

I keep deteriorating. I am sorting out our assets, our land is being sold, I found new homes for our children and am taking care of my husband. [Patient]

It is a strange situation: planning my funeral, writing my obituary, and taking care of other affairs, all while fighting with all of my power to live. [Patient]

Economic Burden

The economic burden of AML or MDS was not frequently mentioned in the analyzed posts. When it was discussed, however, topics related to a lack of income because of the patient's physical inability to work, caregivers not being able to work their standard hours because of additional pressures, and costs of treatment:

As a carer it's very difficult, and sometimes I feel desperate, I have lost my job and now have to manage financial pressures too. [Caregiver]

Discussion

Principal Findings

Patients' perspective is a vital component to understanding treatment decisions patients make and better informing future decisions. The treatment landscape for patients with AML or MDS particularly lacks this frame of reference. Previous research has examined the definition of patients with AML or MDS who are ineligible for intensive chemotherapy, available treatment options, end-of-life care, and health care resource utilization [26-29]. However, no published research, to the best of our knowledge, has investigated patient preferences for treatment of this population or has explored the use of social media via disease-specific forums to research the experiences of these patients and their caregivers. This study was a novel design that combined the use of unstructured social media text and natural language processing with an established text analysis technique, QDA, to attempt to better characterize and understand the perceptions and priorities of patients with AML or MDS who are not eligible for intensive chemotherapy.

Although the focus of this study was patients with AML who are not eligible for intensive chemotherapy, some findings agreed with those of previously published studies that surveyed the more general AML population. The burden of physical impacts of the disease and/or treatment, including fatigue and weakness, infections, and fever, was frequently reported in this study and in published works that examined similar impacts [30,31]. This study also showed how important psychological and emotional impacts (eg, anxiety, depression, fear, and confusion) were to patients. These impacts were often caused by a lack of knowledge of the condition and its treatment or anticipated test results; these aspects were also reflected in previous studies [31].

Social media also provide caregivers and family members with a wider platform to discuss how they are impacted. This study demonstrated the impact on these individuals who were burdened by additional responsibilities to manage the patient's affairs on top of their usual roles and duties. In turn, these burdens sometimes had a negative effect on the emotional and psychological well-being of the caregivers.

The study indicated that patients felt there was a lack of treatment options and did not feel that age alone should be a reason to not receive a particular treatment. There was also a clear theme around a general lack of information and knowledge regarding the condition and treatment options, resulting in the

feeling that the patients and caregivers were unable to make fully informed decisions.

This study provided a unique perspective on the experiences and views of patients dealing with end-of-life situations, a situation not easily explored with a survey-based approach. Patients discussed goals that were broader and more complex than simply extending survival. Patients and caregivers highly valued time at home and with family or desired to reach certain life milestones or family events. These priorities might be considered less by clinicians and researchers, but their impact on mental well-being, survival, and HRQoL can be equally as important as qualitative markers and should be considered during the treatment decision-making process.

A recent survey of patients with AML concluded that treatment options that slow disease progression and do not compromise HRQoL are important in this population [15]. In keeping with this, our targeted analysis found that 44% of patients discussing treatment choices mentioned impact on HRQoL as a key influence in opting for one therapy over another. Surprisingly, aspects related to home and family (42%) were equally influential in choosing a treatment. Patients continued to emphasize the importance of being at home, reaching specific life events, and spending time with family. Interestingly, the negative effect of treatments on HRQoL was most cited in reasons for refusing certain treatments, with patients often stating a preference for higher HRQoL over quantity of life for this condition. Influences under the home and family theme were cited as the reasons for making or accepting certain treatment choices.

The use of social media offers a unique and complementary approach to capture the perspectives of patients and their caregivers. Published studies that aimed to characterize AML patients' perspectives using a patient survey approach varied in sample size from 32 to 82 patients [30-32]. However, these studies looked at a range of AML patient groups and ages and covered signs, symptoms, impacts, positive and negative concepts, receiving information, and making treatment decisions. Buckley et al [30] and LeBlanc et al [32] recruited patients from a single hospital setting, which could have resulted in a lack of more general representation.

This social media study is the first, to our knowledge, to look specifically at the experience of AML and MDS patients ineligible for intensive chemotherapy, and the approach was uniquely adaptable to include more patients as needed. A pool of 347 patients qualified for this study—saturation was reached at 85 for the QDA of objectives other than treatment decisions, and 45 patients were used for a targeted analysis of treatment decision making. The sample sizes in this study were comparable with the largest of the published studies for AML and far exceeded guidelines for sample sizes discussed in the literature [31,33].

The use of social media has the added benefit of providing a more geographically diverse sample and reducing timelines for collecting information. Just as important, if not more so, this approach has the potential to collect unprompted responses by patients and caregivers. Many patients may be too weak or frail to participate in patient surveys and in health-related social media forums, whereas caregivers often participate on the patient's behalf. The inclusion of the experiences of caregivers provides a unique and potentially more comprehensive perspective to the approach of using social media.

A recent publication compared the use of disease-related social media, concept elicitation interviews, and group concept mapping to understand the perspectives of patients with ankylosing spondylitis. The study found that disease-related forums on social media uncovered the most concepts/perceptions, used the least amount of resources, and allowed access to a larger amount of material [34]. Our study also found the use of social media to be an efficient and cost-effective way to collect data of this type. In addition, this study presented a novel insight into the experiences and opinions of a patient group that had not been previously studied. Considering the rarity of the condition and the difficulties in recruiting patients with AML who are not eligible for intensive chemotherapy, this study drew from a relatively large sample size for a more representative look at this patient group. Another strength of this study was that the themes and concepts were driven by the discussions between the patients and caregivers themselves—they were not guided by any moderator, medical professional, or researcher and are therefore less likely to be impacted by information bias than traditional questionnaire or interview studies. Instead, they offer a more accurate and novel reflection of issues that are most important to patients and caregivers.

Social media, however, are a fairly unregulated source of information and should be used with caution. Compared with traditional patient interview studies, the biases present in disease-related social media data are currently not well understood; therefore, it is difficult to characterize the types of bias that could impact the results of this study. However, it should be noted that no study is free of all bias.

Conclusions

Social media provide a window into patients' and caregivers' perspectives using a previously untapped approach. The ability to get a more representative and current depiction of what matters to patients with AML or MDS and their caregivers can have an important influence on the treatment landscape and the resulting decision-making process.

This work highlighted the themes that are important to patients with AML or MDS and their caregivers, which have previously been given little consideration but, nonetheless, have a clear influence on their experiences and treatment decisions. The following themes were considered as very important: improved QoL over extension of life, being at home, spending time with family, and living to reach certain milestones. This study also confirmed what had been previously reported in the more general AML population—physical impacts on QoL such as AEs are important to patients and play a role in treatment decisions. Our approach, however, offered the unique ability to examine a wider range of unsolicited topics and provided a more in-depth look at these discussions.

Physicians and other medical staff should consider these *softer* preferences and goals and use this knowledge to engage in more

comprehensive, informative discussions with patients and caregivers, especially given the lack of treatment options for this patient population. This knowledge also emphasizes the need to close the information gap for patients and caregivers regarding treatment options and can help clinicians make better-informed recommendations for their patients. Furthermore, the findings of this study can be used to inform further research in this population and allow for patients' preferences to be taken into account during drug development and for regulatory decisions.

Acknowledgments

Funding for this research was provided by Pfizer.

Authors' Contributions

All authors provided substantial contributions to the conception and design of work, and interpretation of results and were involved in the review and approval of this manuscript for publication.

Conflicts of Interest

AB, AC, SH, DL, EM, and SP are employees of Evidera (London, UK), who were paid consultants to Pfizer in connection with the development of this manuscript. TB and VW are employees of Pfizer (New York, USA).

Multimedia Appendix 1
Search Terms Used to Identify acute myeloid leukemia (AML) or myelodysplastic syndrome (MDS) specific online forums.
[DOCX File , 27 KB - jmir_v21i11e14285_app1.docx]

References

1. Lubeck D, Danese M, Jennifer D, Miller K, Richhariya A, Garfin PM. Systematic literature review of the global incidence and prevalence of myelodysplastic syndrome and acute myeloid leukemia. Blood 2016;128(22):5930 [FREE Full text]
2. Klepin HD. Myelodysplastic syndromes and acute myeloid leukemia in the elderly. Clin Geriatr Med 2016 Feb;32(1):155-173 [FREE Full text] [doi: 10.1016/j.cger.2015.08.010] [Medline: 26614866]
3. Science Daily. 2017. New Drug Combination Targets Aggressive Blood Cancer. URL: https://www.sciencedaily.com/releases/2017/03/170307114554.htm [accessed 2019-06-28]
4. American Cancer Society. Key Statistics for Acute Myeloid Leukemia (AML). URL: https://www.cancer.org/cancer/acute-myeloid-leukemia/about/key-statistics.html [accessed 2019-04-04] [WebCite Cache ID 77O5ObUgf]
5. American Cancer Society. Key Statistics for Myelodysplastic Syndromes. URL: https://www.cancer.org/cancer/myelodysplastic-syndrome/about/key-statistics.html [accessed 2019-04-04] [WebCite Cache ID 77O5e8j0O]
6. Leukemia & Lymphoma Society. Facts and Statistics. URL: https://www.lls.org/facts-and-statistics/facts-and-statistics-overview/facts-and-statistics [accessed 2019-04-04] [WebCite Cache ID 77O5p5Qla]
7. Thein MS, Ershler WB, Jemal A, Yates JW, Baer MR. Outcome of older patients with acute myeloid leukemia: an analysis of SEER data over 3 decades. Cancer 2013 Aug 1;119(15):2720-2727 [FREE Full text] [doi: 10.1002/cncr.28129] [Medline: 23633441]
8. American Cancer Society. General Approach to Treatment of Myelodysplastic Syndromes. URL: https://www.cancer.org/cancer/myelodysplastic-syndrome/treating/general-approach.html [accessed 2019-04-04] [WebCite Cache ID 77O5xbI3A]
9. Oran B, Weisdorf DJ. Survival for older patients with acute myeloid leukemia: a population-based study. Haematologica 2012 Dec;97(12):1916-1924 [FREE Full text] [doi: 10.3324/haematol.2012.066100] [Medline: 22773600]
10. Ma E, Bonthapally V, Chawla A, Lefebvre P, Swords R, Lafeuille M, et al. An evaluation of treatment patterns and outcomes in elderly patients newly diagnosed with acute myeloid leukemia: a retrospective analysis of electronic medical records from US community oncology practices. Clin Lymphoma Myeloma Leuk 2016 Nov;16(11):625-636.e3 [FREE Full text] [doi: 10.1016/j.clml.2016.08.006] [Medline: 27686689]
11. Medeiros BC, Satram-Hoang S, Hurst D, Hoang KQ, Momin F, Reyes C. Big data analysis of treatment patterns and outcomes among elderly acute myeloid leukemia patients in the United States. Ann Hematol 2015 Jul;94(7):1127-1138 [FREE Full text] [doi: 10.1007/s00277-015-2351-x] [Medline: 25791241]
12. Efficace F, Gaidano G, Breccia M, Criscuolo M, Cottone F, Caocci G, et al. Prevalence, severity and correlates of fatigue in newly diagnosed patients with myelodysplastic syndromes. Br J Haematol 2015 Feb;168(3):361-370. [doi: 10.1111/bjh.13138] [Medline: 25272332]
13. Buckley SA, Kirtane K, Walter RB, Lee SJ, Lyman GH. Patient-reported outcomes in acute myeloid leukemia: Where are we now? Blood Rev 2018 Jan;32(1):81-87. [doi: 10.1016/j.blre.2017.08.010] [Medline: 28888621]
14. Williams LA, Garcia-Gonzalez A, Ahaneku H, Cortes JE, Garcia-Manero G, Kantarjian HM, et al. A patient-reported outcome measure for symptoms of AML/MDS. J Clin Oncol 2015 May;33(Suppl 15):e17648. [doi: 10.1200/jco.2015.33.15_suppl.e17648]

15. Mamolo CM, Cappelleri JC, Hoang CJ, Kim R, Hadfield A, Middleton C, et al. A real-world, cross-sectional, community survey of symptoms and health-related quality of life of adults with acute myeloid leukemia. Future Oncol 2019 Jun;15(16):1895-1909. [doi: 10.2217/fon-2018-0842] [Medline: 30912462]
16. Seo JD, Estey E, Voyard J, O'Donoghue B, Bridges JF. Assessing patient preferences for the benefits and risks of treating for acute myeloid leukemia (AML): a pilot study. Blood 2016;128(22):4799 [FREE Full text]
17. Seo J, Smith BD, Estey E, Voyard E, O' Donoghue B, Bridges JF. Developing an instrument to assess patient preferences for benefits and risks of treating acute myeloid leukemia to promote patient-focused drug development. Curr Med Res Opin 2018 Dec;34(12):2031-2039. [doi: 10.1080/03007995.2018.1456414] [Medline: 29565196]
18. Food and Drug Administration. Patient-Focused Drug Development: Collecting Comprehensive and Representative Input. URL: https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM610442.pdf [accessed 2019-04-04] [WebCite Cache ID 77O69XuHi]
19. Patient Power. Acute Myelogenous Leukemia (AML) Support Groups. URL: https://www.patientpower.info/acute-myeloid-leukemia/support-groups [accessed 2019-04-04] [WebCite Cache ID 77O6GK7W8]
20. Alexa. dailystrength: Overview. URL: https://www.alexa.com/siteinfo/dailystrength.org [accessed 2019-04-04] [WebCite Cache ID 77O6Mo78N]
21. Alexa. macmillan: Overview. URL: https://www.alexa.com/siteinfo/macmillan.org.uk [accessed 2019-04-04] [WebCite Cache ID 77O6Vm0RI]
22. Alexa. mds-foundation: Overview. URL: https://www.alexa.com/siteinfo/mds-foundation.org [accessed 2019-04-04] [WebCite Cache ID 77O6bhGRb]
23. R: The R Project for Statistical Computing. URL: https://www.r-project.org/ [accessed 2019-10-01]
24. CRAN - R Project. 2018. Package 'hunspell'. URL: https://cran.r-project.org/web/packages/hunspell/hunspell.pdf [accessed 2019-04-04] [WebCite Cache ID 77O6iBj8c]
25. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
26. Pettit K, Odenike O. Defining and treating older adults with acute myeloid leukemia who are ineligible for intensive therapies. Front Oncol 2015;5:280 [FREE Full text] [doi: 10.3389/fonc.2015.00280] [Medline: 26697412]
27. Estey E. Therapeutic options for patients who are not eligible for intensive chemotherapy. Mediterr J Hematol Infect Dis 2013;5(1):e2013050 [FREE Full text] [doi: 10.4084/MJHID.2013.050] [Medline: 23936621]
28. Lübbert M, Suciu S, Baila L, Rüter BH, Platzbecker U, Giagounidis A, et al. Low-dose decitabine versus best supportive care in elderly patients with intermediate- or high-risk myelodysplastic syndrome (MDS) ineligible for intensive chemotherapy: final results of the randomized phase III study of the European Organisation for Research and Treatment of Cancer Leukemia Group and the German MDS Study Group. J Clin Oncol 2011 May 20;29(15):1987-1996. [doi: 10.1200/JCO.2010.30.9245] [Medline: 21483003]
29. Erba HP. Finding the optimal combination therapy for the treatment of newly diagnosed AML in older patients unfit for intensive therapy. Leuk Res 2015 Feb;39(2):183-191 [FREE Full text] [doi: 10.1016/j.leukres.2014.11.027] [Medline: 25577399]
30. Buckley SA, Jimenez-Sahagun D, Othus M, Walter RB, Lee SJ. Quality of life from the perspective of the patient with acute myeloid leukemia. Cancer 2018 Jan 1;124(1):145-152 [FREE Full text] [doi: 10.1002/cncr.30982] [Medline: 28881384]
31. Tomaszewski EL, Fickley CE, Maddux L, Krupnick R, Bahceci E, Paty J, et al. The patient perspective on living with acute myeloid leukemia. Oncol Ther 2016;4(2):225-238 [FREE Full text] [doi: 10.1007/s40487-016-0029-8] [Medline: 28261652]
32. LeBlanc TW, Fish LJ, Bloom CT, El-Jawahri A, Davis DM, Locke SC, et al. Patient experiences of acute myeloid leukemia: a qualitative study about diagnosis, illness understanding, and treatment decision-making. Psychooncology 2017 Dec;26(12):2063-2068. [doi: 10.1002/pon.4309] [Medline: 27862591]
33. Mason M. Sample size and saturation in PhD studies using qualitative interviews. Qual Soc Res 2010;11(3): Article 8. [doi: 10.17169/fqs-11.3.1428]
34. Humphrey L, Willgoss T, Trigg A, Meysner S, Kane M, Dickinson S, et al. A comparison of three methods to generate a conceptual understanding of a disease based on the patients' perspective. J Patient Rep Outcomes 2017;1(1): Article 9 [FREE Full text] [doi: 10.1186/s41687-017-0013-6] [Medline: 29757313]

Abbreviations

AE: adverse event
AML: acute myeloid leukemia
HRQoL: health-related quality of life
MDS: myelodysplastic syndrome
QDA: qualitative data analysis
QoL: quality of life

Edited by N Bruining; submitted 09.04.19; peer-reviewed by DM Zhang, C Wang; comments to author 20.06.19; revised version received 01.08.19; accepted 23.08.19; published 22.11.19.

Please cite as:
Booth A, Bell T, Halhol S, Pan S, Welch V, Merinopoulou E, Lambrelli D, Cox A
Using Social Media to Uncover Treatment Experiences and Decisions in Patients With Acute Myeloid Leukemia or Myelodysplastic Syndrome Who Are Ineligible for Intensive Chemotherapy: Patient-Centric Qualitative Data Analysis
J Med Internet Res 2019;21(11):e14285
URL: http://www.jmir.org/2019/11/e14285/
doi:10.2196/14285
PMID:31755871

Original Paper

Assessing the Appeal of Instagram Electronic Cigarette Refill Liquid Promotions and Warnings Among Young Adults: Mixed Methods Focus Group Study

Linnea I Laestadius[1], MPP, PhD; Kendall E Penndorf[1], MPH; Melissa Seidl[1], MPH; Young I Cho[1], PhD

Zilber School of Public Health, University of Wisconsin-Milwaukee, Milwaukee, WI, United States

Corresponding Author:
Linnea I Laestadius, MPP, PhD
Zilber School of Public Health
University of Wisconsin-Milwaukee
PO Box 413
Milwaukee, WI, 53201-0413
United States
Phone: 1 414 227 4512
Email: llaestad@uwm.edu

Abstract

Background: While marketing for electronic cigarette refill liquids (e-liquids) is widespread on Instagram, little is known about the post elements that create appeal among young adult Instagram users. Further information is needed to help shape regulatory strategies appropriate for social media.

Objective: This study examined young adult Instagram user perceptions of actual e-liquid marketing posts and US Food and Drug Administration (FDA)–mandated nicotine addiction warning statements on Instagram.

Methods: A series of 12 focus groups (n=69) were held with non–tobacco users, vapers, smokers, and dual users in Wisconsin between September and December 2018. Participants discussed the elements of posts that they found appealing or unappealing, in addition to completing a survey about each post and e-liquid. Focus group transcripts were analyzed by smoking status using a framework analysis approach.

Results: Although willingness to try e-liquids was highest among nicotine users, focus group discussions indicated that Instagram posts promoting e-liquids held appeal for individuals across smoking statuses. The primary elements that created appeal were the perceived trustworthiness of the Instagram account, attractive design and flavor visuals, and promotion of flavors and nicotine levels that met personal preferences. Post appeal was reduced by references to vaping subcultures, indicators that the post creator did not take nicotine addiction seriously, and FDA-mandated nicotine warning statements. Non–tobacco users were particularly drawn to posts featuring nicotine-free e-liquids with attractive visual designs and flavors known from foods.

Conclusions: Young adults consider a broad range of elements in assessing the appeal of e-liquid marketing on Instagram, with minor but notable distinctions by smoking status. Non–tobacco users are uniquely drawn to nicotine-free e-liquids and are more deterred by the FDA's mandated nicotine addiction warning statements than those from other smoking statuses. This suggests that it may be possible to tailor policy interventions in a manner that helps to reduce novel uptake of vaping without significantly diminishing its potential harm-reduction benefits.

(J Med Internet Res 2019;21(11):e15441) doi:10.2196/15441

KEYWORDS

social media; vaping; tobacco; marketing

Introduction

With the growing popularity of refillable electronic cigarettes (e-cigarettes) [1], marketing for the aerosolized electronic cigarette refill liquids (e-liquids) used with devices has become increasingly prevalent. Early estimates suggested that there were thousands of flavors of e-liquids available for sale online [2]. Considering the extremely large variety of e-liquids on the market, brands have sought to distinguish their products using creative flavors, names, label designs, and advertising visuals [3,4]. Many e-liquids are promoted on the visual social media platform Instagram, which is used by 71% of young adults aged

18 to 24 in the United States [5]. While Instagram bans paid advertising for tobacco products, both e-liquids and e-cigarettes are regularly promoted through posts made by vendors, brands, and sponsored users [3,6,7]. This content remains largely unregulated, except for US Food and Drug Administration (FDA) regulations requiring nicotine addiction warning statements to be placed on all visual advertising for e-liquids containing nicotine. The FDA has clarified that these provisions apply to advertising "communicated via mobile telephone, smartphone, microblog, social media Web site, or other communication tool" [8].

Despite the rapid growth of this content, little is known about the elements of promotional e-liquid posts that shape appeal among young adults. Analyses of Instagram post metadata suggest that e-cigarette posts focused on products receive more likes than explicitly promotional posts [7]. Experimental studies using artificial posts indicate that celebrity endorsements of e-cigarettes on Instagram are associated with greater intent to use e-cigarettes [9,10], although formal celebrity vaping promotions are not currently prevalent on Instagram [3,11]. Exposure to cartoons on e-liquid bottles has also been shown to be associated with a susceptibility to use e-cigarettes [12]. Studies of Instagram promotion from other fields suggest that high quality images and avoiding resemblances to traditional advertising are key to establishing appeal and authenticity [13,14]. There is also limited evidence on the ways in which FDA-mandated nicotine warnings might alter the appeal of posts. To date, only one study has considered these warnings on social media. Focusing on a sample of e-cigarette users, the study found that warnings on tweets from a fictional e-cigarette company reduced perceptions of healthiness but did not increase perceptions of harm or reduce willingness to try e-cigarettes [15].

To address current gaps in the literature and determine the post elements that shape appeal, this study examined perceptions of actual Instagram posts that promote e-liquids through a series of twelve focus groups with young adults. Evidence suggests that young adults are at higher risk of initiating e-cigarette use than youth [16], making this a particularly important population for study. Considering both the importance of e-cigarettes as a harm reduction tool for smokers and the challenges of growing uptake of e-cigarettes by nonsmokers [17], we also considered how perspectives differ by smoking status. Finally, we examined the impact of FDA-mandated warning statements about nicotine addiction. Study insights can help to determine priorities for the regulation of e-cigarette promotions on visual social media.

Methods

Summary

Young adults in the Milwaukee, Wisconsin area were recruited through ads on Facebook, Instagram, and Craigslist, as well as flyers distributed at local vape and tobacco stores, coffee shops, libraries, and one public and one private university. To be eligible, participants had to be aged 18-24 years old, speak English, and view their Instagram account at least once per week. Participants were screened by a graduate research assistant and those who met the above criteria were invited to join the study. Participants were assigned into focus groups based on their status as a smoker, vaper, dual user, or non–tobacco user. Former smokers who did not currently vape were excluded from enrollment. Participants were provided with a meal and a $50 gift card as an incentive. The Institutional Review Board at the University of Wisconsin-Milwaukee approved the research protocol.

Data Collection

Overall, 12 focus groups were held between September and December 2018. Prior research suggests that most themes are identified in the first three focus groups, so we aimed for three groups per smoking status [18]. Focus groups were stratified by smoking status and sizes ranged from four to eight participants. Each focus group lasted approximately 90 minutes. A semistructured interview guide was developed to elicit perceptions of five real world Instagram posts promoting e-liquids. Posts were chosen to be representative of common themes, user types, and practices found in e-liquid promotions on Instagram [3]. Specifically, posts were chosen to reflect both vendors and individual users with commercial ties to brands. Posts were also chosen to reflect elements such as warning statements, flavor descriptions and visuals, cartoon illustrations, nicotine and nicotine-free e-liquids, and the presence of a person vaping. To protect the privacy of the accounts whose posts were included, Multimedia Appendix 1 includes a written description of posts rather than post reproductions. To capture greater variation in post elements and include more posts with the FDA nicotine warning statements, which became mandatory effective August 2018, three of the posts were switched out for new ones after the completion of the first focus group for each smoking status. Mandated warnings comprise 20% of the visual element of the post, and state: "WARNING: This product contains nicotine. Nicotine is an addictive chemical."

Following informed consent, participants completed questionnaires on self-described demographics, tobacco use, and prior e-liquid marketing exposure. Participants were also presented with printed copies of each post. To provide participants with the full context needed to assess posts, these were paired with images of the user profile pages associated with posts (containing the user's profile information and small versions of their three most recent Instagram posts). For each post, participants were asked to circle elements they found appealing or unappealing and complete a survey indicating if they would like, comment on, or repost the post, their willingness to purchase or try the e-liquid, and their perception of the e-liquid using a response scale, modeled after prior research on perceptions of e-cigarette ads [19] and based on the following word pairs: delicious/disgusting, enjoyable/unenjoyable, healthy/unhealthy, safe/dangerous, fun/boring, and cool/not cool. For each word pair, participants used a 5-point scale to indicate which word better captured their impressions (eg, fun would be a 5 and boring would be a 1). Led by a professional, independent moderator, each post was discussed one at a time, with participants prompted to explain elements that they found appealing or unappealing. All focus groups were audio-recorded.

Data Analysis

First, quantitative data from the survey were analyzed across smoking status groups. Because each individual participant rated 5 Instagram images and therefore those 5 ratings are nonindependent, meaning that ratings are clustered within each individual (a total of 345 ratings by 69 participants), analyzing such nonindependent data as if they were independent can lead to biased estimates. To account for nonindependence, we employed mixed-effects models (mixed-effects logistic models for binary variables, and mixed-effects regression models for the 5-point Likert scale values) [20]. All statistical analyses were conducted using STATA 14 (StataCorp, College Station, Texas).

Following transcription of focus group discussions, qualitative data were analyzed using a framework analysis approach [21]. The first transcript from each smoking status was independently coded line-by-line by the lead (LL), second (KP), and third (MS) authors to allow for discussion of commonly found themes and the creation of an initial codebook. Themes were identified using an inductive process, with initial codes being grouped into higher level codes through team discussion [22]. Following this, the codebook was applied to all transcripts by the first author (LL) and second author (KP) using MAXQDA 2018 (VERBI Software, Berlin, Germany). All coding was discussed, with any disagreements resolved and updates to code definitions made as needed. Following coding, all data were charted by primary theme and smoking status into a spreadsheet matrix to facilitate analysis and interpretation. Additionally, memos were used to summarize and explore relations between themes, identify differences by smoking status, and highlight illustrative quotes. Memos took the form of multiple collaborative Word documents exploring different themes to assist with data immersion and the leap from "the concrete to the conceptual" [23]. The findings below capture themes related to the elements that shape the appeal of e-liquid marketing on Instagram. For brevity, smokers, vapers, and dual users together are collectively referred to as "nicotine users".

Results

Overview

We conducted a total of 12 focus groups with 69 individuals; three focus groups each with vapers (n=20), dual users (n=18), smokers (n=12), and non–tobacco users (n=19). Overall, the sample was 58% male (n=40) and 41% female (n=28), with an average age of 21.10. One participant identified as nonbinary. See Table 1 for further participant demographics. A total of 45% (n=9) of vapers were former smokers, while 58% (n=7) of smokers reported having tried e-cigarettes at least once. Overall, 50% (n=9) of dual users and 83% (n=10) of smokers reported smoking cessation attempts in the past year. Post surveys indicated that about one third of participants would like the posts if they saw them on their Instagram feed. Only 7% (n=5) and 4% (n=3) of them would comment or repost, respectively. The group difference on these opinions was found not to be statistically significant. However, vapers were more likely to rate e-liquids "healthy" and "safe" than nonusers. All nicotine user statuses also expressed a greater willingness to try the e-liquids if offered by a close friend and to buy them for themselves. See Table 2 for full post ratings and Table 3 for full e-liquid ratings. In Table 3, means represent ratings from 1-5, with 5 representing the most favorable rating of the e-liquid. Standard errors were adjusted for clustering of the measures within respondents.

Table 1. Participant characteristics by smoking status.

Characteristics	Total (N=69)	Nonuser (n=19)	Cigarette smoker (n=12)	Vaper (n=20)	Dual user (n=18)
Age (mean, SD)	21.10 (1.91)	21.11 (2.08)	22.92 (1.31)	19.75 (1.21)	21.39 (1.65)
Gender, n (%)					
Male	40 (58.0)	9 (47.4)	9 (75.0)	14 (70.0)	8 (44.4)
Female	28 (40.6)	9 (47.4)	3 (25.0)	6 (30.0)	10 (55.6)
Nonbinary	1 (1.4)	1 (5.2)	0 (0)	0 (0)	0 (0)
Race/ethnicity, n (%)					
White	43 (62.3)	9 (47.4)	3 (25.0)	17 (85.0)	14 (77.8)
African American/black	11 (15.9)	6 (31.6)	3 (25.0)	0 (0)	2 (11.1)
Hispanic/Latinx	3 (4.4)	0 (0)	3 (25.0)	0 (0)	0 (0)
Asian	6 (8.7)	1 (5.3)	2 (16.7)	1 (5.0)	2 (11.1)
American Indian	1 (1.5)	1 (5.3)	0 (0)	0 (0)	0 (0)
Biracial	5 (7.2)	2 (10.5)	1 (8.3)	2 (10.0)	0 (0)
Education, n (%)[a]					
High school graduate	9 (13.2)	1 (5.6)	1 (8.3)	4 (20.0)	3 (16.7)
Some college or technical training	47 (69.1)	11 (61.1)	8 (66.7)	16 (80.0)	12 (66.7)
College graduate or more	12 (17.6)	6 (33.3)	3 (25.0)	0 (0)	3 (16.7)
Seen/heard promos for e-cigs[b]/e-liquids[c], n (%)[a]					
In stores	45 (66.2)	13 (72.2)	10 (83.3)	11 (55.0)	11 (61.1)
TV	11 (16.2)	7 (38.9)	0 (0)	2 (10.0)	2 (11.1)
Radio	2 (2.9)	1 (5.6)	1 (8.3)	0 (0)	0 (0)
Instagram	40 (58.8)	12 (66.7)	6 (50.0)	12 (60.0)	10 (55.6)
Other social media	40 (58.8)	13 (72.2)	5 (41.7)	11 (55.0)	11 (61.1)
Website banner ads	20 (29.4)	7 (38.9)	1 (8.3)	8 (40.0)	4 (22.2)
Billboards	7 (10.3)	2 (11.1)	2 (16.7)	2 (10.0)	1 (5.6)

[a]n=18. One nonuser did not list education or prior marketing exposure.

[b]e-cigs: electronic cigarettes.

[c]e-liquids: electronic cigarette refill liquids.

Table 2. Ratings of the posts by smoking status.

Ratings	Total, proportion (SD)	Nonuser, proportion (SD)	Cigarette smoker, proportion (SD)	*P* value[a]	Vaper, proportion (SD)	*P* value[a]	Dual user, proportion (SD)	*P* value[a]
Would like post	0.33 (0.04)	0.28 (0.06)	0.48 (0.10)	.06	0.33 (0.06)	.61	0.30 (0.07)	.93
Would comment on post	0.07 (0.02)	0.13 (0.06)	0.08 (0.04)	.75	0.05 (0.03)	.28	0.02 (0.02)	.11
Would repost	0.04 (0.01)	0.01 (0.01)	0.08 (0.05)	.13	0.02 (0.02)	.79	0.05 (0.03)	.27

[a]*P* values were obtained from mixed-effect models with nonusers as a reference group.

Table 3. Ratings of the electronic cigarette refill liquids by smoking status.

Ratings	Total, mean (SD)	Nonuser, mean (SD)	Cigarette smoker, mean (SD)	*P* value[a]	Vaper, mean (SD)	*P* value[a]	Dual user, mean (SD)	*P* value[a]
Delicious	3.21 (0.08)	3.03 (0.15)	3.17 (0.21)	.48	3.19 (0.13)	.45	3.41 (0.15)	.08
Enjoyable	3.31 (0.08)	3.10 (0.15)	3.27 (0.21)	.53	3.37 (0.16)	.23	3.48 (0.15)	.10
Healthy	2.51 (0.10)	1.95 (0.14)	2.51 (0.15)	.05	3.04 (0.19)	<.001	2.47 (0.20)	.04
Safe	2.85 (0.10)	2.37 (0.13)	2.90 (0.22)	.07	3.30 (0.21)	<.001	2.78 (0.19)	.11
Fun	3.44 (0.09)	3.13 (0.14)	3.64 (0.32)	.07	3.46 (0.16)	.17	3.58 (0.13)	.07
Cool	3.04 (0.12)	2.62 (0.20)	3.22 (0.35)	.09	3.14 (0.17)	.08	3.21 (0.21)	.05
Would try	3.61 (0.12)	2.80 (0.29)	3.54 (0.21)	.02	4.01 (0.12)	<.001	3.97 (0.21)	<.001
Would buy	2.15 (0.10)	1.52 (0.15)	2.24 (0.23)	.007	2.26 (0.16)	.003	2.57 (0.21)	<.001

[a]*P* values were obtained from mixed-effect models with nonusers as a reference group.

Focus Group Discussions

Overview

Through framework analysis we identified the following themes in how young adults assess the appeal of e-liquid marketing posts on Instagram: (1) accounts must be trustworthy; (2) visuals are key to grabbing attention; (3) flavor mimicry, in which flavors resemble known foods and beverages, is favored; (4) nicotine levels should be clear and have options; (5) references to vaping culture repel users and nonusers alike; (6) marketing should be sensitive to nicotine addiction; and (7) warning statements reduce post appeal. Each of these and their subthemes are covered below, with discussion of any differences tied to smoking status. Illustrative quotes are found in Table 4.

Accounts Must be Trustworthy

Participants of all smoking statuses frequently reflected on the importance of the trustworthiness of the users who made the posts. Limited trust in a user, or the company they were promoting, diminished interest in trying a product due to both safety and ethical concerns. Without the ability to try a product in person, participants relied on proxy measures for quality. The number of followers, likes, and comments were frequently mentioned as a good sign that a poster was trustworthy (Table 4: Q1.1). Several participants were concerned about the ratios of followers to following and about fake followers, which caused trust to significantly diminish. Across all the groups, many participants found posts from individual Instagram users with brand/store sponsorships to be appealing and authentic and therefore more trustworthy than more explicitly commercial posts made directly by brands or stores (Table 4: Q1.2). The more the post felt like organic, unpaid content, the more trust increased. For many there was a sense of inherent trust in the posters, and sponsorship conferred some sense of the product having been vetted. One non–tobacco user felt that sponsored posts were particularly reliable because:

> *you're not going to just like let yourself be sponsored by anything.*

However, when sponsored users crossed the line into being "someone who just has sponsors for money," their higher levels of appeal begin to diminish (Table 4: Q1.3).

Table 4. Selected quotes highlighting key themes in responses to e-liquid[a] promotion posts.

Theme	Quote[b]
Accounts must be trustworthy and authentic.	
1.1	*The more [followers] you have the more accountable you are to more people, so I trust something from someone who has a lot more followers than someone who has a lot less.* [V[c] 47]
1.2	*I just feel more trust for this post than I would someone that's like purely trying to sell me something... It's just more of a personal picture because it could be one of my friends posting a picture of what they're smoking rather than like a graphic that someone obviously spent money for a designer [to create]…* [DU[d] 26]
1.3	*Like maybe he's not a company, but he's definitely like a promoter and someone is paying him to do this… it takes me out of the whole personal thing, it doesn't feel as personal.* [DU 32)]
Visuals are key to grabbing attention.	
2.1	*Regardless of the how good the flavors sound, you're not going to look in the description unless you like the main picture...* [DU 20]
2.2	*I'm not going to lie, I marked like it looked fun mostly because of the colors within it...* [NU[e] 2]
2.3	*…it's easy to look at that and I think of the flavor the strawberry and imagining like how it kind of does make you think of gum and like you can easily imagine what a fruity minty gum tastes like and so, even though it's not gum obviously, but it's easy to image what that would be like tasting it and also I smoke menthol cigarettes so, the whole- the specific thing of like refreshing, menthol, and then ice cubes it's like "yes, that's- I like that, that's up my alley".* [CS[f] 21]
Flavor mimicry is favored.	
3.1	*If I didn't use it, this would be the one that would make me want to, simply because I like candy and it's something that I know I would like if it actually did imitate the taste or something like that, so I thought it was cool.* [NU 21]
3.2	*Yeah, I mean I love coffee so it definitely stood out as intriguing, caught my attention.* [NU 26]
3.3	*Because it doesn't have nicotine and it looks like a pie flavor so, I have smoked regular cigarettes and it's not like you have these kinds of flavors. It's just something that I may want to explore... It's something that I might want to explore, you know, to try different flavors like in a smoking experience.* [CS 23]
Nicotine levels should be clear and have options.	
4.1	*I hit mine just for the buzz. I don't really see the point of hitting for a cloud [of vapor]… I wouldn't buy it because I don't really see the point.* [V 47]
4.2	*I know the difference between lights and Marlboro lights and Marlboro reds and menthol and that kind of stuff but, if someone tells me like, "Okay, this vape has 21 milligrams of nicotine," I have no idea what that amounts to....So, instead of just putting out numbers make it easier in a way to understand people that look at this as, this has more nicotine as compared to a normal one. This is something which has less.* [CS 27]
4.3	*If I were to vape, I would probably start with something that doesn't have nicotine and, how am I supposed to know, if it had like a label that's probably what I'm going to reach for.* [NU 09]
References to vaping culture repel users and nonusers alike	
5.1	*It makes me not want to buy it like at all... Yeah. I don't know, it's just overly masculine. Yeah "be like us use this e-liquid" or whatever, just like some Chad. I don't know, I hate everything about this ad.* [V 31]
5.2	*There's no imagery of like vaping in movies and shit…you don't see like a Swashbucklers, like a cool character, or like bad boy with a vape. It's always a cigarette.* [CS 18]
5.3	*None of [the posts] seem like they are at all geared towards quitting smoking, it's just promoting the hobbyist side of vaping...I don't know, if I were to go through the process of quitting something, I wouldn't want to make it this glamorized thing of like "oh now I'm joining this cool vaping community." I'd just be like "oh wow, I'm damaging my body every day and I want to try to damage my body a little bit less." I don't know if I would necessarily buy into that as "this is my new hobby now," like it's the same as if someone said that their hobby was going sober and not drinking alcohol...* [DU 20]
Marketing should be sensitive to nicotine addiction	
6.1	*I feel like it's representing the vape community poorly and makes it seem like we're trying to market, or they're trying to market, to young people and there are just a bunch of average people like us who vape, for whatever reason that we do, and aren't trying to promote this negatively. So, I think it just reflects poorly.* [V 30]
6.2	*The whole point of e-cigarettes is to help people use a healthier alternative to smoking and eventually quit nicotine or cigarettes all together, but this guy is treating it as a hobby which is like super lame. For example, like the zero milligrams and he even has a hashtag that says "stop smoking." Like there's really no point of him even putting this in his lungs if he's smoking zero milligrams.* [CS 18]

Theme	Quote[b]
6.3	*You shouldn't brag about being addicted to something that isn't good for you. Like it's- not negative it just doesn't need to be there, like if there is like some alcohol company showing off their bottle, I guarantee they wouldn't put 'alcoholic' as a hashtag.* [V 29]
Warning statements reduce post appeal	
7.1	*It's kind of like album artwork when they have the explicit sign on it and it's really pretty album artwork it's like – I get it, it has swearwords, why did you have to put that there? You just ruined the album artwork. And it's kind of the same idea.* [V 40]
7.2	*I already smoke cigarettes, there's already nicotine in that so-(laughs).* [CS 4]
7.3	*I don't know, it's like putting 'warning' on like a beer thing, "warning this beer will get you drunk," like yeah, no shit, it will.* [V 31]

[a]e-liquid: electronic cigarette refill liquids.

[b]Quotes are identified by smoking status (V, DU, CS, or NU) and participant number.

[c]V: vaper.

[d]DU: dual user.

[e]NU: non–tobacco user.

[f]CS: cigarette smoker.

Visuals are Key to Grabbing Attention

All participant groups favored good visual design, defined largely by how well posts met conventions for "good Instagramming" (eg, not too many hashtags, not reposting identical images, use of aesthetically appealing filters), the clarity with which posts conveyed flavors and nicotine levels, and personal preferences for colors, visuals, and themes used in marketing. Given the Instagram ecosystem, in which users must choose to stop on a post and read it rather than continuing to scroll through their feed, some participants explicitly recognized the importance of initial post impression (Table 4: Q2.1).

Although thematic and aesthetic preferences were highly individual, with participants in the same group often strongly divided on the visual appeal of a post, most users favored fun and bright-colored posts (Table 4: Q2.2). Pop culture themes were also attractive to many. A post featuring an e-liquid with an anthropomorphic cartoon banana on the label and the caption "This flavor is B-A-N-A-N-A-S" was particularly popular and made several participants think of a song by pop artist Gwen Stefani. Overly edgy themes were more broadly disliked due to heightened risk perceptions and a sense of companies trying too hard to be cool. Visuals depicting flavors were almost universally appreciated when the foods aligned with their own taste preferences. Flavor visuals allowed participants, including non–tobacco users, to conjure up a vivid image of the flavor (Table 4: Q2.3).

Flavor Mimicry is Favored

As suggested above by the importance of flavor visuals, the flavors of the e-liquids being marketed were critical for determining appeal. Although individual flavor preferences were varied, which suggests that multiple flavor options are critical, most participants expressed liking flavors that mimicked tastes that they already knew and enjoyed in the context of real foods and beverages. Several non–tobacco users expressed strong interest in known flavors (Table 4: Q3.1, Q3.2). Specific named items like branded candies were particularly appealing because they were considered to be increasing the odds of flavors tasting good.

Unlike non–tobacco users, who mainly felt that flavors seemed appealing, vapers, dual users, and smokers had concerns with flavor fidelity and the extent to which flavors accurately captured the flavor they were intended to portray. Several e-liquids were deemed unappealing because the users felt that manufacturers probably failed to capture the flavor correctly. Vague e-liquid names and posts without clear flavor information also reduced post appeal. Many vapers and dual users felt that visuals were important for initial appeal, but flavor was what that ultimately shaped their perception of the e-liquids. As a result of this, several participants indicated that they would be willing to try e-liquids in the posts but would not outright buy them in case they disliked the flavor. For smokers who did not currently vape, e-liquid flavors presented a distinct point of interest because it was a way to branch out from the taste of tobacco. Some smokers indicated that they might entertain the idea of vaping because of the flavors depicted in the posts (Table 4: Q3.3)

Nicotine Levels Should be Clear and Have Options

Participants from all smoking statuses wanted posts to provide clear information about nicotine levels. Among nicotine users, there was a desire to know about the availability of higher nicotine levels than the ones depicted in the post. This was particularly the case for posts depicting bottles with 0 to 3 milligrams of nicotine, which many nicotine users found to be insufficient to meet their own needs (Table 4: Q4.1). Still, most vapers and dual users felt more inconvenienced than deterred by unappealing nicotine levels because they indicated that the e-liquid flavor would likely be sold in multiple nicotine levels if they looked it up online. Smokers seemed less aware of the variety of nicotine levels available and were not always clear on which level of nicotine would meet their needs (Table 4: Q4.2). Overall, nicotine users also had an appreciation for e-liquids offered in a variety of nicotine levels, with several noting that nicotine users could gradually taper their nicotine levels for health or cessation.

Non–nicotine users wanted to know which e-liquids were nicotine-free and considered clear labeling to be useful. Some suggested this could be like peanut-free or sugar-free labeling, which was partially out of concern for former smokers who they envisioned might want to vape but not use nicotine, while others appeared to consider their own potential use of e-liquids (Table 4: Q4.3). The presence of nicotine was seen as elevating the risks of the product, while non–nicotine products were more welcoming and less risky.

References to Vaping Culture Repels Users and Nonusers Alike

Participants felt that several of the Instagram posts captured images, themes, and language that were representative of vaping culture. The presence of these post elements reduced appeal among large numbers of participants across all smoking status groups. A post depicting a tattooed man in a baseball cap vaping and holding an e-liquid elicited particularly negative responses. The frequent use of vape culture hashtags referencing women (eg, #girlswhovape, #dripgirls), despite the lack of depiction of women in posts, was particularly unappealing to female participants. Several of the hashtags used, such as #vapeordie and #swag, were seen as "cringey" and off-putting. Despite their own e-cigarette use, vapers and dual users were also extremely put-off by post elements referencing their visions of stereotypical vaping culture (Table 4: Q5.1). Non–tobacco users and smokers both suggested that the posts were trying too hard to be appealing and that vaping was less cool than smoking (Table 4: Q5.2).

Negative responses to marketing elements drawing on vaping culture were grounded in a preexisting, extremely negative view of stereotypical vapers. Across the smoking statuses, stereotypical vapers were commonly made fun of as "douches," "Chads," "neck-beards," and "bros." Several smokers and non–tobacco users suggested that vaping was inherently less cool than smoking. Participants from all smoking statuses were reluctant to be associated with vaping culture. Vapers and dual users spoke of trying to disassociate themselves from the culture depicted in the posts, while smokers were concerned that vaping would automatically lump them into the vaping culture. One smoker felt that the post sent a message that "if you vape, you will become like this." The poor perception of vaping culture appeared to serve as a deterrent for some dual users and smokers to switch entirely to vaping (Table 4: Q5.3).

Marketing Should be Sensitive to Nicotine Addiction

Across all smoking statuses, several participants were upset about the ways in which post creators failed to consider the seriousness of nicotine addiction. Many users recognized that youth vaping was an issue and several complained that posts and e-liquids contained elements, such as sweet flavors and cartoons, that might be too appealing to younger users. Some of these concerns were grounded in concern for youth, while others were concerned that it made the vaping community look bad (Table 4: Q6.1). Two common concerns specific to nicotine users were that posts trivialized addiction and that they failed to promote products in a way that met the needs of those with nicotine addictions (Table 4: Q6.2). Nicotine users were generally accepting of their own addictions but expressed significant concern about the broader ramifications of marketing strategies that targeted children or would lead to continued addiction. Hashtags such as #vapeaddict created particularly negative responses (Table 4: Q6.3).

Warning Statements Reduce Post Appeal

Several participants across smoking statuses felt that the FDA-required warning statements about nicotine being addictive reduced post appeal. However, the impact on risk perception and appeal of the products promoted in posts was more ambiguous. Primarily, the warnings were seen as diminishing appeal because they were so large that they ruined the aesthetics and visual design of the post (Table 4: Q7.1). Some questioned the judgement of the Instagram user for applying such large and frequent warnings, not realizing that they were mandated. Other participants were aware of the requirement but were unclear about when the statements were required, particularly regarding sponsored users and e-liquids without nicotine. At the same time that they expressed dislike over the visual impact of the warnings, many nicotine users felt that the presence of the warnings was important to deter youth and novel users, although they were not always sure that the warnings would actually work to deter anyone.

Regarding personal risk perception, nicotine users were largely nonplussed by the warnings about nicotine addiction due to their current nicotine addictions and product familiarity (Table 4: Q7.2, Q7.3). Others felt that warning statements were so omnipresent in their lives, particularly on foods, that they tuned them out altogether. However, a few nicotine users did explicitly suggest that it was an unwelcome reminder of the potential harms of their vaping. For non–tobacco users, the warning statements had a clearer impact on risk perception and reduced product appeal. Many saw themselves as non–nicotine users, as reflected in the higher appeal of zero milligram nicotine e-liquids. Accordingly, the presence of the nicotine warnings reduced product appeal, and products in posts without warnings were often seen as safer. One non–tobacco user explained that the warnings were like a "road closed do not go here sign."

Discussion

Primary Findings

The findings from this focus group study suggest that Instagram posts promoting e-liquids may hold appeal for individuals across smoking statuses, including non–tobacco users. The primary elements that create appeal are the trustworthiness of the Instagram account, the use of attractive design and flavor visuals, and the promotion of flavors and nicotine levels that meet personal preferences. The importance of visuals, particularly those depicting flavors, suggests that Instagram posts are well positioned for highlighting the often detailed and vivid art on e-liquid bottles. By contrast, post appeal is reduced by references to vaping subcultures, themes and hashtags that suggest post creators do not take nicotine addiction seriously, and the use of FDA-mandated warning statements. In short, participants weighed the account creator, the post content, and the product being promoted in making their assessment of posts. Each of these represent an area for potential regulatory action.

The elevated appeal and sense of authenticity created by sponsored users is of concern since Instagram is home to large volumes of posts made by users who indicate a sponsor or affiliate relationship with e-liquid and e-cigarette brands [3]. Prior Instagram research also supports that ads that resemble user-generated posts are considered more authentic and appealing than those that resemble traditional advertising [14,24]. Given the spread and appeal of sponsored promotions for e-liquids, this represents an important area for future research and action by both the FDA and the US Federal Trade Commission.

While findings suggest that limiting flavor options would likely help reduce post appeal among nonsmokers, it would likely have a similar effect on smokers and dual users who may, in the absence of full cessation, benefit from a full transition to e-cigarette use. Several prior studies have also suggested that non–tobacco flavors may play a role in smoking cessation for young adults [25,26]. Regulating the use of colors and cute/cartoon visuals on labels, which were a significant driver of initial appeal and a source of concern for some nicotine users, may be a more pragmatic approach. Further research should explore the viability of limiting appeal specifically among non–tobacco users. FDA-mandated warning statements on images do appear to have a promising effect on reducing the appeal of posts. Attractive and platform normative visuals play an essential role to establishing marketing appeal on Instagram [13,14], and it appears that large warnings may disrupt the positive effects of an otherwise good visual. This impact is distinct from the actual risk messaging of the warning, which had a more limited negative effect that was largely limited to nonsmokers. Echoing findings from Guillory et al [15], the actual content of the warnings may not be effective for nicotine users given that they already experience nicotine addiction. Further, regulators should ensure that the warning mandate does not inadvertently lead to more posts marketing nicotine-free e-liquids to avoid applying the warnings. Focus group findings suggest that an increase in marketing for zero milligram e-liquids may cause increased interest from non–nicotine users, while having minimal effects on e-cigarette users.

Finally, references to vaping subculture appeared to be the element that most strongly reduced the appeal of posts among participants. While early studies suggested that young adults found vaping to be “cool” [27,28], more recent work suggests that vaping has shifted into a stigmatized “uncool” behavior [29-32]. In this case, it appears that one of the most unappealing things an e-liquid post can do is prominently feature an image of a person who is seen as a stereotypical vaper. This is a particularly notable finding given earlier research suggesting that images of a person using an e-cigarette created appeal among smokers [33,34]. Both the drivers of this shift and the implications for e-cigarette uptake need further study. The health equity risks of stigmatizing e-cigarette users in a similar fashion to smokers should also be considered [35].

Limitations

This study has several limitations. Although focus group participants were recruited from multiple social media platforms and offline settings, they may not be representative of all young adults. The overall number of smokers enrolled was also smaller than other status groups due to a low prevalence of smokers among those who contacted the study team to enroll. Additionally, smoking status was self-reported, and some participants may have been placed in focus groups not reflecting their true smoking status. The use of real Instagram posts and brands prevented a narrow focus on the impact of individual post elements in isolation, but significantly strengthened ecological validity.

Conclusion

This research suggests that young adults consider a broad range of elements in assessing the appeal of e-liquid marketing on Instagram. Although flavors are important for ultimate appeal, participants also mentioned factors such as visual design, nicotine levels, and account trustworthiness as critical. The intersection between Instagram, as a visual platform, and the often-elaborate art on e-liquid bottles is particularly notable for those seeking to reduce the appeal of e-liquid marketing. For young adults who did not currently use tobacco, posts featuring attractive e-liquid designs in multiple flavors and zero milligrams of nicotine were particularly appealing and therefore of concern. Warnings reduced the visual appeal of posts among participants, although the specific phrasing mandated by the FDA may have limited impact on those who already face addictions to nicotine. Overall, the identification of nuances in post perceptions across smoking status groups suggests that it may be possible to target regulatory approaches specifically toward minimizing vaping appeal among non–tobacco users.

Acknowledgments

This work was supported by the National Cancer Institute at the National Institutes of Health and the FDA Center for Tobacco Products (grant number R03CA216528). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the FDA. The authors would like to thank Alana Schneider, Alex Fairburn, Holly Milaeger, and Cameron Macdonald for their contributions to recruitment, planning, and facilitation, and Pallav Pokhrel for early feedback.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Summary of posts used to guide focus group discussion.
[DOCX File , 16 KB - jmir_v21i11e15441_app1.docx]

References

1. Barrington-Trimis JL, Gibson LA, Halpern-Felsher B, Harrell MB, Kong G, Krishnan-Sarin S, et al. Type of E-Cigarette Device Used Among Adolescents and Young Adults: Findings From a Pooled Analysis of Eight Studies of 2166 Vapers. Nicotine Tob Res 2018 Jan 05;20(2):271-274 [FREE Full text] [doi: 10.1093/ntr/ntx069] [Medline: 28371890]
2. Zhu S, Sun JY, Bonnevie E, Cummins SE, Gamst A, Yin L, et al. Four hundred and sixty brands of e-cigarettes and counting: implications for product regulation. Tob Control 2014 Jul;23 Suppl 3:iii3-iii9 [FREE Full text] [doi: 10.1136/tobaccocontrol-2014-051670] [Medline: 24935895]
3. Laestadius LI, Wahl MM, Pokhrel P, Cho YI. From Apple to Werewolf: A content analysis of marketing for e-liquids on Instagram. Addict Behav 2019 Apr;91:119-127. [doi: 10.1016/j.addbeh.2018.09.008] [Medline: 30253933]
4. Allem J, Cruz TB, Unger JB, Toruno R, Herrera J, Kirkpatrick MG. Return of cartoon to market e-cigarette-related products. Tob Control 2019 Sep;28(5):555-557 [FREE Full text] [doi: 10.1136/tobaccocontrol-2018-054437] [Medline: 30049688]
5. Smith A, Anderson M. Social Media Use 2018: Demographics and statistics.: Pew Research Center: Internet, Science & Tech; 2018 Mar 1. URL: https://www.pewinternet.org/2018/03/01/social-media-use-in-2018/ [accessed 2019-05-30]
6. Majmundar A, Kirkpatrick M, Cruz TB, Unger JB, Allem J. Characterising KandyPens-related posts to Instagram: implications for nicotine and cannabis use. Tob Control 2019 May 30:1. [doi: 10.1136/tobaccocontrol-2019-055006] [Medline: 31147484]
7. Chu K, Allem J, Cruz TB, Unger JB. Vaping on Instagram: cloud chasing, hand checks and product placement. Tob Control 2016 Sep;26(5):575-578 [FREE Full text] [doi: 10.1136/tobaccocontrol-2016-053052] [Medline: 27660111]
8. U.S. Food and Drug Administration. 81 FR 28973. 2016 May 10. Deeming Tobacco Products to be Subject to the Federal Food, Drug, and Cosmetic Act, as Amended by the Family Smoking Prevention and Tobacco Control Act URL: https://www.federalregister.gov/documents/2016/05/10/2016-10685/deeming-tobacco-products-to-be-subject-to-the-federal-food-drug-and-cosmetic-act-as-amended-by-the [accessed 2019-10-01]
9. Phua J, Jin SV, Hahm JM. Celebrity-endorsed e-cigarette brand Instagram advertisements: Effects on young adults' attitudes towards e-cigarettes and smoking intentions. J Health Psychol 2018 Mar;23(4):550-560. [doi: 10.1177/1359105317693912] [Medline: 28810409]
10. Phua J, Lin J, Lim DJ. Understanding consumer engagement with celebrity-endorsed E-Cigarette advertising on instagram. Computers in Human Behavior 2018 Jul;84(5):93-102. [doi: 10.1016/j.chb.2018.02.031]
11. Lee AS, Hart JL, Sears CG, Walker KL, Siu A, Smith C. A picture is worth a thousand words: Electronic cigarette content on Instagram and Pinterest. Tob Prev Cessat 2017 Jul;3:1-9 [FREE Full text] [doi: 10.18332/tpc/74709] [Medline: 28815224]
12. Kirkpatrick MG, Cruz TB, Unger JB, Herrera J, Schiff S, Allem J. Cartoon-based e-cigarette marketing: Associations with susceptibility to use and perceived expectations of use. Drug Alcohol Depend 2019 Aug 01;201:109-114. [doi: 10.1016/j.drugalcdep.2019.04.018] [Medline: 31207451]
13. Teo LX, Leng HK, Phua YXP. Marketing on Instagram. International Journal of Sports Marketing and Sponsorship 2019 May 07;20(2):321-332. [doi: 10.1108/IJSMS-04-2018-0028]
14. Colliander J, Marder B. 'Snap happy' brands: Increasing publicity effectiveness through a snapshot aesthetic when marketing a brand on Instagram. Computers in Human Behavior 2018 Jan;78:34-43. [doi: 10.1016/j.chb.2017.09.015]
15. Guillory J, Kim AE, Fiacco L, Cress M, Pepper J, Nonnemaker J. An experimental study of nicotine warning statements in ecigarette tweets. Nicotine Tob Res 2019 Mar 01:ntz029. [doi: 10.1093/ntr/ntz029] [Medline: 30820571]
16. Perry CL, Pérez A, Bluestein M, Garza N, Obinwa U, Jackson C, et al. Youth or Young Adults: Which Group Is at Highest Risk for Tobacco Use Onset? J Adolesc Health 2018 Oct;63(4):413-420 [FREE Full text] [doi: 10.1016/j.jadohealth.2018.04.011] [Medline: 30001826]
17. Warner KE, Mendez D. E-cigarettes: Comparing the Possible Risks of Increasing Smoking Initiation with the Potential Benefits of Increasing Smoking Cessation. Nicotine Tob Res 2019 Jan 01;21(1):41-47. [doi: 10.1093/ntr/nty062] [Medline: 29617887]
18. Guest G, Namey E, McKenna K. How Many Focus Groups Are Enough? Building an Evidence Base for Nonprobability Sample Sizes. Field Methods 2016 Jul 24;29(1):3-22. [doi: 10.1177/1525822X16639015]
19. Duke JC, Allen JA, Eggers ME, Nonnemaker J, Farrelly MC. Exploring Differences in Youth Perceptions of the Effectiveness of Electronic Cigarette Television Advertisements. Nicotine Tob Res 2016 May;18(5):1382-1386. [doi: 10.1093/ntr/ntv264] [Medline: 26706908]
20. Rabe-Hesketh S, Skrondal A. Multilevel and Longitudinal Modeling Using Stata, 3rd ed. College Station, TX: Stata Press; 2012.
21. Gale NK, Heath G, Cameron E, Rashid S, Redwood S. Using the framework method for the analysis of qualitative data in multi-disciplinary health research. BMC Med Res Methodol 2013 Sep 18;13:117 [FREE Full text] [doi: 10.1186/1471-2288-13-117] [Medline: 24047204]
22. Berends L, Johnston J. Using multiple coders to enhance qualitative analysis: The case of interviews with consumers of drug treatment. Addiction Research & Theory 2009 Jul 11;13(4):373-381. [doi: 10.1080/16066350500102237]

23. Birks M, Chapman Y, Francis K. Memoing in qualitative research. Journal of Research in Nursing 2008 Jan;13(1):68-75. [doi: 10.1177/1744987107081254]
24. Johnson B, Potocki B, Veldhuis J. Is that my friend or an advert? The effectiveness of Instagram native advertisements posing as social posts. Journal of Computer-Mediated Communication 2019 May;24(3):108-125 [FREE Full text] [doi: 10.1093/jcmc/zmz003]
25. Chen JC. Flavored E-cigarette Use and Cigarette Smoking Reduction and Cessation-A Large National Study among Young Adult Smokers. Subst Use Misuse 2018 Oct 15;53(12):2017-2031. [doi: 10.1080/10826084.2018.1455704] [Medline: 29624135]
26. Harrell MB, Weaver SR, Loukas A, Creamer M, Marti CN, Jackson CD, et al. Flavored e-cigarette use: Characterizing youth, young adult, and adult users. Prev Med Rep 2017 Mar;5:33-40 [FREE Full text] [doi: 10.1016/j.pmedr.2016.11.001] [Medline: 27896041]
27. Kong G, Morean ME, Cavallo DA, Camenga DR, Krishnan-Sarin S. Reasons for Electronic Cigarette Experimentation and Discontinuation Among Adolescents and Young Adults. Nicotine Tob Res 2015 Jul;17(7):847-854 [FREE Full text] [doi: 10.1093/ntr/ntu257] [Medline: 25481917]
28. Coleman BN, Johnson SE, Tessman GK, Tworek C, Alexander J, Dickinson DM, et al. "It's not smoke. It's not tar. It's not 4000 chemicals. Case closed": Exploring attitudes, beliefs, and perceived social norms of e-cigarette use among adult users. Drug Alcohol Depend 2016 Feb 01;159:80-85 [FREE Full text] [doi: 10.1016/j.drugalcdep.2015.11.028] [Medline: 26708706]
29. Martinez LS, Hughes S, Walsh-Buhi ER, Tsou M. "Okay, We Get It. You Vape": An Analysis of Geocoded Content, Context, and Sentiment regarding E-Cigarettes on Twitter. J Health Commun 2018;23(6):550-562. [doi: 10.1080/10810730.2018.1493057] [Medline: 29979920]
30. Katz SJ, Erkinnen M, Lindgren B, Hatsukami D. Beliefs about E-cigarettes: A Focus Group Study with College Students. Am J Health Behav 2019 Jan 01;43(1):76-87. [doi: 10.5993/AJHB.43.1.7] [Medline: 30522568]
31. Robertson L, Hoek J, Blank M, Richards R, Ling P, Popova L. Dual use of electronic nicotine delivery systems (ENDS) and smoked tobacco: a qualitative analysis. Tob Control 2019 Jan;28(1):13-19 [FREE Full text] [doi: 10.1136/tobaccocontrol-2017-054070] [Medline: 29419488]
32. Tokle R, Pedersen W. "Cloud chasers" and "substitutes": e-cigarettes, vaping subcultures and vaper identities. Sociol Health Illn 2019 Jun;41(5):917-932. [doi: 10.1111/1467-9566.12854] [Medline: 30677161]
33. Maloney EK, Cappella JN. Does Vaping in E-Cigarette Advertisements Affect Tobacco Smoking Urge, Intentions, and Perceptions in Daily, Intermittent, and Former Smokers? Health Commun 2016;31(1):129-138. [doi: 10.1080/10410236.2014.993496] [Medline: 25758192]
34. Pepper JK, Emery SL, Ribisl KM, Southwell BG, Brewer NT. Effects of advertisements on smokers' interest in trying e-cigarettes: the roles of product comparison and visual cues. Tob Control 2014 Jul;23 Suppl 3:iii31-iii36 [FREE Full text] [doi: 10.1136/tobaccocontrol-2014-051718] [Medline: 24935896]
35. Burgess DJ, Fu SS, van Ryn M. Potential unintended consequences of tobacco-control policies on mothers who smoke: a review of the literature. Am J Prev Med 2009 Aug;37(2 Suppl):S151-S158. [doi: 10.1016/j.amepre.2009.05.006] [Medline: 19591755]

Abbreviations

e-cigarettes: electronic cigarettes
e-liquids: electronic cigarette refill liquids
FDA: United States Food and Drug Administration

Edited by G Eysenbach; submitted 10.07.19; peer-reviewed by S Katz, JP Allem, M Anderson, Z Zhang; comments to author 23.09.19; revised version received 26.09.19; accepted 27.09.19; published 25.11.19.

Please cite as:
Laestadius LI, Penndorf KE, Seidl M, Cho YI
Assessing the Appeal of Instagram Electronic Cigarette Refill Liquid Promotions and Warnings Among Young Adults: Mixed Methods Focus Group Study
J Med Internet Res 2019;21(11):e15441
URL: http://www.jmir.org/2019/11/e15441/
doi:10.2196/15441
PMID:31763987

Original Paper

Exploring Research Priorities of Parents Who Have Children With Down Syndrome, Cleft Lip With or Without Cleft Palate, Congenital Heart Defects, or Spina Bifida Using ConnectEpeople: A Social Media Coproduction Research Study

Marlene Sinclair[1], BSc, MEd, PGDipEd, RM, PhD; Julie EM McCullough[1], BSc, PGDip, PhD; David Elliott[2], BSc, MBA, CPhys, PhD; Anna Latos-Bielenska[3], MD, PhD; Paula Braz[4], MSc; Clara Cavero-Carbonell[5], BSPharm, MPH, PhD; Anna Jamry-Dziurla[3], MSc; Ana João Santos[4,6], PhD; Lucía Páramo-Rodríguez[5], BA, MA

[1]Institute of Nursing and Health Research, Ulster University, Newtownabbey, Northern Ireland, United Kingdom

[2]Redburn Solution Ltd, Belfast, United Kingdom

[3]Department of Medical Genetics, Poznan University of Medical Sciences, Poznan, Poland

[4]Epidemiology Department, National Institute of Health Doctor Ricardo Jorge, Lisbon, Portugal

[5]Rare Diseases Research Unit, Foundation for the Promotion of Health and Biomedical Research in the Valencian Region, Valencia, Spain

[6]Public Health Research Centre, National School of Public Health, Nova University Lisbon, Lisbon, Portugal

Corresponding Author:
Marlene Sinclair, BSc, MEd, PGDipEd, RM, PhD
Institute of Nursing and Health Research
Ulster University
Shore Road, Room 12J09
Newtownabbey, Northern Ireland, BT37 0QB
United Kingdom
Phone: 44 02890368118
Email: m.sinclair1@ulster.ac.uk

Abstract

Background: Using social media for research purposes is novel and challenging in terms of recruitment, participant knowledge about the research process, and ethical issues. This paper provides insight into the recruitment of European parents of children with specific congenital anomalies to engage in coproduction research by using social media. Secret Facebook groups, providing optimal security, were set up for newly recruited research-aware parents (RAPs) to communicate privately and confidentially with each other and for the research team to generate questions and to interpret findings.

Objective: This study aimed to use social media for the recruitment and engagement of parents in research and to determine the research priorities of parents who have children with Down syndrome, cleft lip with or without cleft palate, congenital heart defects, and spina bifida.

Methods: The design was exploratory and descriptive with 3 phases. Phase 1 included the recruitment of RAPs and generation of research questions important to them; phase 2 was a Web-based survey, designed using Qualtrics software, and phase 3 included analysis and ranking of the top 10 research questions using an adapted James Lind Alliance approach. Simple descriptive statistics were used for analysis, and ethical approval was obtained from the Ethics Filter Committee of the Institute of Nursing and Health Research, Ulster University.

Results: The recruitment of 32 RAPs was a sensitive process, varying in the time taken to consent (mean 51 days). However, parents valued the screening approach using the State-Trait Anxiety Inventory as a measure to ensure their well-being (mean 32.5). In phase 1, RAPs generated 98 research questions. In phase 2, 251 respondents accessed the Web-based survey, 248 consented, and 80 completed the survey, giving a completeness rate of 32.3% (80/248). Most parents used social media (74/80, 92%). Social media, online forums, and meeting in person were ranked the most preferable methods for communication with support groups networks and charities. Most respondents stated that they had a good understanding of research reports (71/80, 89%) and statistics (68/80, 85%) and could differentiate among the different types of research methodologies (62/80, 78%). Phase 3 demonstrated consensus among RAPs and survey respondents, with a need to know the facts about their child's condition, future health, and psychosocial and educational outcomes for children with similar issues.

Conclusions: Social media is a valuable facilitator in the coproduction of research between parents and researchers. From a theoretical perspective, ocularcentrism can be an applicable frame of reference for understanding how people favor visual contact.

(J Med Internet Res 2019;21(11):e15847) doi:10.2196/15847

KEYWORDS

e-forum; social media; Web-based survey; Facebook; STAI; Down syndrome; cleft lip with or without cleft palate; congenital heart defects; spina bifida; parents; ocularcentrism; coproduction

Introduction

Background

The European Commission [1] highlights the value of using social media for communication and engagement with the public and acknowledges that it is a "beneficial tool to connect with others" and to "find new research partners." Parents require health-related information about their children [2], and an increasing number seek this information from the internet and on social media platforms [3-5], with many going to online forums to discuss specific issues [6]. This is particularly true for parents who have a child with a chronic health condition [7]. Parents strongly feel that they can and must have a voice in health and education research that will have a positive impact on their child's everyday life [8,9]. Across Europe, awareness of the benefits of patient and public involvement (PPI) in health care research is rapidly increasing [10]. Lander et al [11] has identified a number of benefits of actively engaging with service users, including the development of research goals that are congruent with those of the public and "assessing the impact and value of health technologies and health services." However, there remains wide variation among countries in the opportunities to do so [10]. In the United Kingdom, PPI in research refers to researchers and patients, carers, and the public working in collaborative partnership to add value to the research process in an accessible and meaningful way [12]. The UK National Institute for Health Research (NIHR) incorporates and funds INVOLVE [12] to "support active public involvement in the National Health Service, public health and social care research." "INVOLVE defines public involvement in research as research being carried out 'with' or 'by' members of the public rather than 'to,' 'about' or 'for' them" [12].

Social media platforms can provide the basis for reciprocal, real-time discussion and sharing of valuable, high-quality information among members with a strong *human-to-human connection* [13]. Russell et al [14] developed a study to connect parents participating in research and researchers in Canada. Using a secret Facebook group, an online discussion network was developed. The group was built by and for parents of children with special needs working with researchers to develop relevant research questions and priorities. The families involved stated that it was a valuable resource for social support and sourcing information in a secure and private way. They felt that the platform provided them with an opportunity to have their voice heard and this was empowering. Researchers involved in the study reported that they were able to connect and discuss with parents directly, which gave them a clearer understanding of the daily life and struggles of families who have a child with special needs [14].

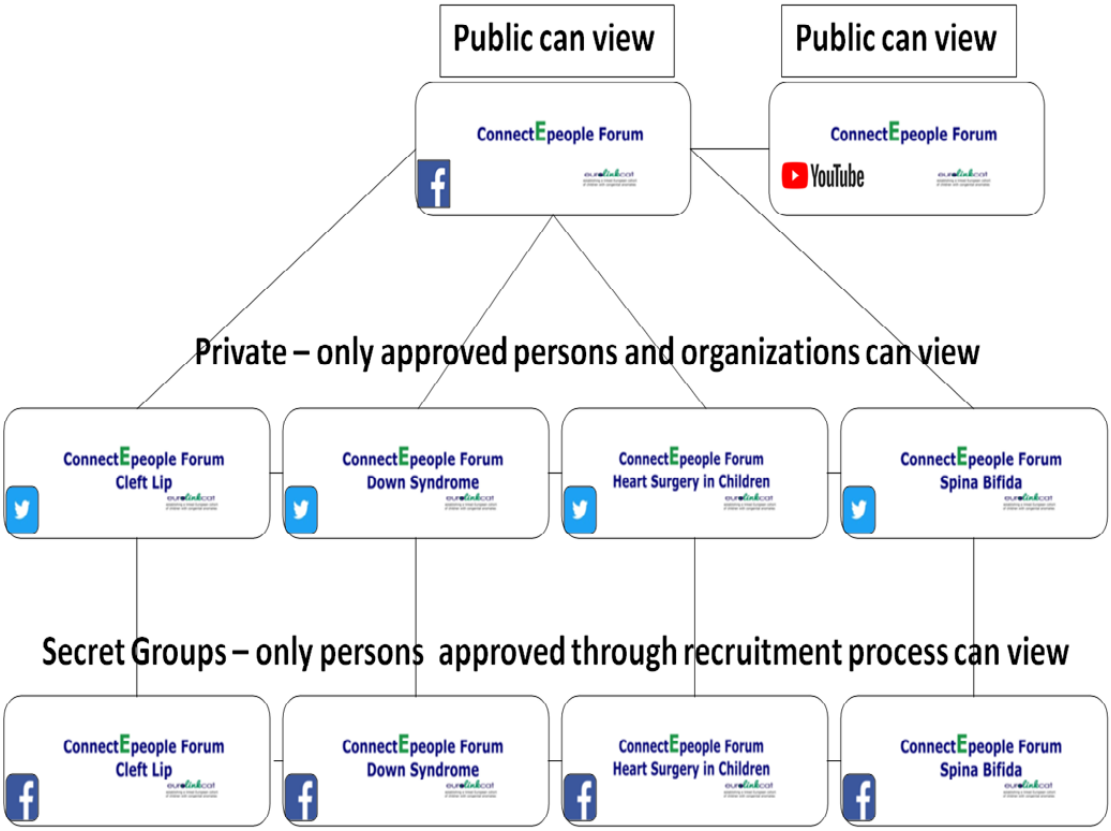
From a theoretical perspective, understanding how and why people value technology for its combination of *electronic touching* and instantaneous access to *visible data* is an important determinant in their preference for the use of social media [15]. The term *ocularcentrism* has become familiar to researchers in social media and is a phenomenon that is built on the theory that *seeing* is believing [16]. Messages, communicated using technology that is embedded in social media with illustrative and graphical sophistication, optimize the visual representation of data in three dimensions, word, text, and video, making it more powerful, believable, and acceptable.

Objectives

Establishing a linked European Cohort of Children with Congenital Anomalies (EUROlinkCAT) is a European project with a number of aims, one of which is connecting researchers and families of children with specific congenital anomalies (CAs), such as Down syndrome (DS), cleft lip with or without cleft palate (CLP), congenital heart defects (CHD), and spina bifida (SB), under the banner of ConnectEpeople. The aim of this project was to actively involve parents in setting research priorities and ensuring that research results are disseminated in a meaningful way by establishing a sustainable electronic forum (e-forum), called ConnectEpeople, to provide regional, national, and international support to families through maintaining the links between the European Surveillance of Congenital Anomalies' (EUROCAT) registries and families [17]. Therefore, this online forum was designed to maximize public and professional engagement in research by establishing a public Facebook page, 4 private Twitter accounts, a YouTube channel, and 4 secret Facebook groups as a cohesive moderated platform: the ConnectEpeople e-forum (Figure 1).

The aim of ConnectEpeople was to use social media for recruitment and engagement of parents in research and to determine the research priorities of parents who have children with DS, CHD, CLP, and SB.

Figure 1. The ConnectEpeople electronic forum structure.



Methods

Overview

The design was exploratory and descriptive and used a mixed methods approach in 3 phases. Parents who had a child with 1 of the 4 CAs were actively recruited via social media to participate in an online research forum using secret Facebook groups. Participating parents were known as research-aware parents (RAPs). Working in partnership with RAPs, an online introduction to the overall project objectives of EUROlinkCAT [18] was made available, and participants were asked to familiarize themselves with the content at the beginning of the study. RAPs provided the research team with questions that were important to them in relation to their child's condition. A Web-based survey was developed, which was open to any parent of a child with 1 of the 4 conditions. The findings from the survey were analyzed and discussed with the RAPs for further feedback and clarification.

Study Phases

Phase 1: Parent Engagement and the Generation of Research Questions

Parent Engagement and Recruitment

Parent support organizations across 9 European countries identified by the research team were contacted [19]. Those who agreed to act as gatekeepers provided information about the ConnectEpeople project to their members via their social media profiles, via group newsletters, in person, on the telephone, and via other communication networks. The researcher contacted the parents, provided them with the participant information sheet, and arranged a time to discuss details about the project and the parents' research needs face-to-face using Skype, WhatsApp, or FaceTime, via video chat on Facebook Messenger, or meeting in person.

Parents living in Croatia, France, Germany, Italy, Netherlands, Poland, Portugal, Spain, and the United Kingdom interested in joining the ConnectEpeople project were invited to do so if they met the following inclusion criteria: they were a parent of a child with CLP, DS, CHD, or SB; a member of a recognized CA parent organization; their child was aged between 1 and 11 years; they had access to social media; and they were able to understand English. Potential participants were screened to ensure they were in good psychological health to participate in the study. The screening procedure involved a self-completion measure of their current anxiety level by completing the State-Trait Anxiety Inventory (STAI) [20] online. Various reliability and validity tests have been conducted on the STAI and have provided sufficient evidence that it is an appropriate, reliable, and adequate measure for studying anxiety in research and clinical settings [21] and indicates anxiety levels for a single point in time. The average STAI scores for working adult (aged 19-69 years) females is 35.2 (SD 10.61) and males is 35.72 (SD 10.40); the score is also dependent on the age of the respondent [20]. In a South American validation study of the STAI and the Beck Depression Inventory, the mean STAI scores for anxious participants was 52.8 (SD 11.4) [22]. Therefore, given the nature of the medical conditions under consideration, STAI anxiety scores above the average for this population were expected, and the cutoff score was set at 65. Once parents completed the STAI,

which was available in 13 languages, and received their score, they were asked to sign a consent form and a project-specific social media policy that outlined the standards ConnectEpeople required participants to observe when using social media. Following this, parents were invited to join the ConnectEpeople secret Facebook group. Secret Facebook groups are not visible to the public, and membership was by invitation only from the moderators (MS/JMcC). In the group, communication was restricted to the specific CA cohort, thus increasing confidentiality and encouraging open dialogue. Recruitment ran from January 2018 to March 2019.

Generation of Research Questions

Research questions for inclusion in the Web-based survey were identified with RAPs using Facebook, WhatsApp, Skype, and video chat, during the ConnectEpeople recruitment process. RAPs used email, telephone, and discussion in secret Facebook groups to identify questions. During the discussions with RAPs, the researcher provided details about the EUROlinkCAT project's aims and specific research objectives to collect data on *education*, *morbidity*, and *survival* and the role of ConnectEpeople. All RAPs were asked the open question "Is there a research question that you already have that you would like an answer to?" Questions generated through this process were compiled and reviewed by 2 researchers for inclusion or exclusion in the project survey.

Phase 2: Survey Development

A Web-based survey was developed using Qualtrics software. The survey was designed in English and translated into Polish, Portuguese, and Spanish. The translation process included changes necessary to ensure that the survey was culturally, socially, and regionally acceptable to the respondents. The survey was reviewed for face and content validity with 12 researchers with experience in using Web-based surveys and/or CAs and 7 RAPs, all of whom were based in Poland, Portugal, or the United Kingdom. Following this pilot stage, minor amendments were made. The duration and the complexity of the survey were reduced by applying survey logic, and thus, only the questions that were relevant to each respondent, based on previous answers, were displayed. The incorporation of a back button allowed respondents to change their answers.

The Web-based survey comprised 62 items, and duplicate entries were avoided by preventing users' access to the survey twice. In addition to seeking verification on research questions generated by RAPs in phase 1, on a 4-point Likert scale of *really important*, *important*, *not sure*, or *definitely not important*, additional questions were focused on obtaining information on demographics, modes of communication with support groups, research knowledge, and use of the internet for research.

Respondents were asked to rate questions on the use of the internet for research-related searches on a 5-point Likert scale of *strongly agree*, *somewhat agree*, *neither agree nor disagree*, *somewhat disagree*, and *strongly disagree*. The survey finished with an open request for parents to identify a question on their research needs related to their child's conditions.

All survey data and personal data were stored in 1 password-protected Qualtrics account and an appropriate password-protected computer on a password-protected network within the Institute of Nursing and Health Research, Ulster University, United Kingdom.

Phase 3: Data Analysis of Survey

Following the closure of the survey, all data were cleaned, checked for errors, and analyzed within the Qualtrics system using simple descriptive statistics. For analysis purposes, the Likert items were recoded into 3 categories; research questions generated by RAPs that were reported as *really important* and *important* were combined. Similarly, for the questions related to research needs, responses of *strongly agree* and *somewhat agree* were combined and *strongly disagree* and *somewhat disagree* were combined. Following the identification of the most important research questions, the RAPs in the 4 ConnectEpeople secret Facebook groups were consulted to seek consensus on relevancy and ranking.

The James Lind Alliance (JLA) [23] is a research initiative that brings patients, carers, and clinicians together in "Priority Setting Partnerships (PSPs) to identify and prioritize the Top 10 unanswered questions or evidence uncertainties that they agree are the most important" [23]. The overall aim was to ensure that researchers and funding bodies understand the key health questions that the patients and the public view as research priorities. Therefore, drawing from the JLA guidelines, the research team sought to use an adapted approach to develop a list of parent's top 10 research priorities. Survey respondents' rating of the proposed research questions were reviewed and confirmed for relevancy by the RAPs in their secret Facebook groups.

Ethical Considerations

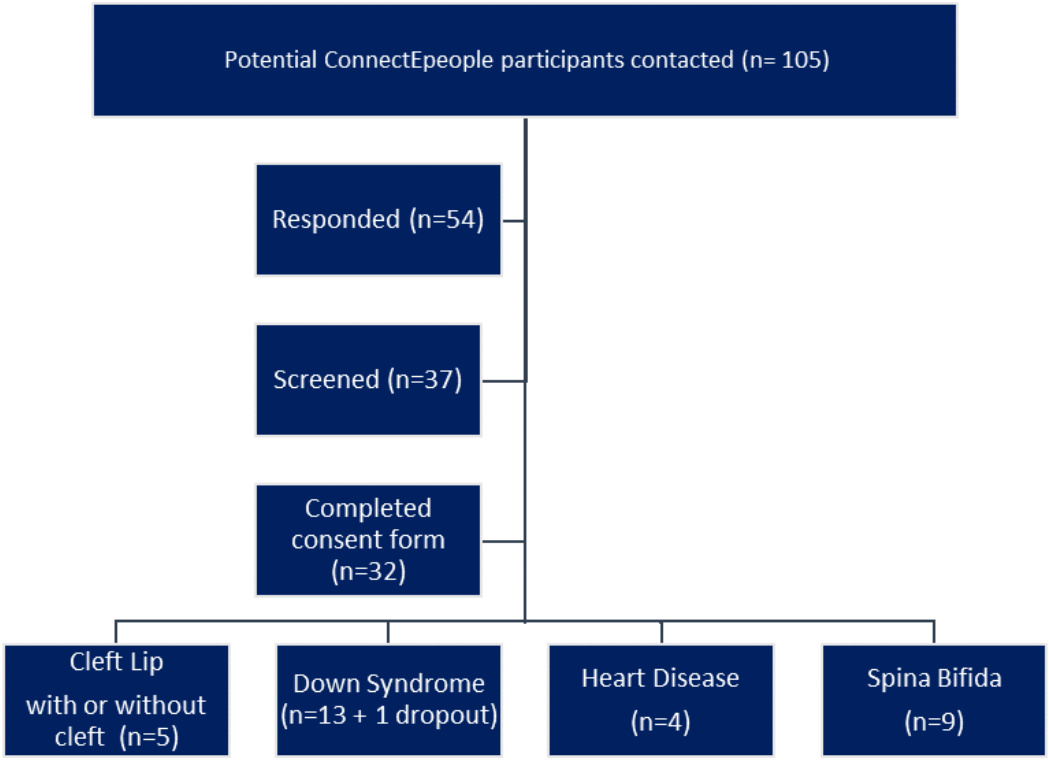
Ethical approval for the study was obtained from the Ethics Filter Committee of the Institute of Nursing and Health Research, Ulster University, on November 21, 2017.

Results

Research-Aware Parents

Recruitment took place online for all participants and was welcomed by parents who found working online was easier for them. Following the notification of interest in ConnectEpeople, 105 parents were contacted, of whom 54/105 (51.4%) responded, 38/105 (36.2%) completed the screening process, and 32/105 (30.5%) entered the secret Facebook element of the ConnectEpeople e-forum (Figure 2). RAPs came from 7 countries: Croatia, Germany, Italy, Poland, Portugal, Spain, and the United Kingdom. The average duration for recruitment from first contact by the parent until entry into the secret Facebook groups was 51 days.

Figure 2. Recruitment of research-aware parents to the ConnectEpeople project.



The inclusion of the STAI for screening purposes demonstrated to parents that the ConnectEpeople research team cared about their health and well-being:

> *…no one ever cared about my mental health before.* [RAP, DS, Poland]

The mean STAI scores across all 4 CAs were below the cutoff score identified for use in this project, and all parents who were screened were invited to take part. The average STAI score across all 4 groups was 32.5; the lowest scores were for RAPs whose children had CLP, and the highest scores were for those with a child with CHD (Multimedia Appendix 1). No parent was excluded based on their STAI score.

A total of 98 questions were identified through conversations with RAPs for inclusion in the survey; 28 questions were generated by RAPs who had a child with DS, 23 questions by RAPs who had a child with CLP, 22 questions by RAPs who had a child with SB, and 25 questions by RAPs who had a child with CHD. In the secret Facebook groups, RAPs discussed the survey respondents' rating of the questions. Rating of research questions demonstrated a high level of consistency between RAPs and survey respondents for all 4 CA groups.

Survey Respondents

A total of 251 parents accessed the survey, 248 consented to take part, and 227 completed the first page of the survey. Respondents were from the following 17 countries: Bulgaria, Croatia, Germany, India, Ireland, Lithuania, Netherlands, Panama, Peru, Poland, Portugal, South Africa, Spain, Turkey, the United Arab Emirates, the United Kingdom, and the United States. Overall, 100 partially completed surveys from parents of a child with DS (35/100), SB (39/100), CL*P* (16/100), and CHD (10/100) provided sufficient data for analysis. Moreover, 80 respondents fully completed the survey, giving a completeness rate of 32.3% (80/248). Table 1 illustrates the demographic profile and use of social media.

Table 1. Demographic details of ConnectEpeople survey respondents who fully completed the survey and their use of social media (N=80).

Characteristic	Value
Parent, n (%)	
Mother	75 (94)
Father	5 (6)
Age (years), mean (SD)	38 (6.77)
Relationship status, n (%)	
With partner	74 (93)
No partner	6 (7)
Member of a parent support organization, n (%)	
Yes	66 (83)
No	14 (17)
Use of social media, n (%)	
Facebook	74 (93)
WhatsApp	66 (83)
Twitter	20 (25)
Snapchat	16 (20)
Instagram	6 (8)
Able to speak English, n (%)	
Yes	79 (99)
No	1 (1)
Educational attainment, n (%)	
Secondary education	4 (5)
Diploma	14 (18)
Undergraduate	29 (36)
Postgraduate	33 (41)
Condition of child, n (%)	
Down syndrome	28 (35)
Spina bifida	28 (35)
Cleft lip with or without cleft palate	16 (20)
Congenital heart defects	8 (10)
Country of residence, n	
United Kingdom	32 (40)
Portugal	14 (18)
Poland	12 (15)
Ireland	9 (11)
Germany	2 (2.5)
Netherlands	2 (2.5)
Spain	2 (2.5)
Bulgaria	1 (1)
Croatia	1 (1)
India	1 (1)
Lithuania	1 (1)
Panama	1 (1)

Characteristic	Value
Turkey	1 (1)
United States	1 (1)

Use of Social Media and the Internet

In total, 73/80 (91%) respondents used a number of methods to communicate with support groups; Multimedia Appendix 2 demonstrates their communication preferences. Social media was preferred because of its accessibility, speed of contact, visual choices, ease of use, multiple links, and 24-hour availability. Mobile phones were the main method of connecting to the internet (62/80, 78%).

Face-to-face discussion was convenient and preferred for personal and confidential conversations, and discussion forums were the most popular because of the opportunity for personal sharing among people who have similar issues of concern. This was supported by comments from respondents and RAPs:

> *Using technology and social media is the way forward and the best way to get parents involved as practically it can be difficult to have the time to meet face to face or at particular times, this is more flexible.* [RAP, SB, United Kingdom]

> *It gives the broader range of possibilities, you can not only read and watch but you can contribute, you can share links.* [Respondent, DS, Poland]

Most respondents felt that it was important for parents to be able to understand basic research (71/80, 89%). The majority had a good understanding of research reports (71/80, 89%) and statistics (68/80, 85%) and could differentiate between different types of research methodologies (62/80, 78%).

Research Questions

Tables 2-5 show the top 10 questions for each CA ranked by importance based on the survey results. Further thematic analysis resulted in the identification of the following 4 main subthemes:

- Facts: All parents were concerned about the facts concerning their child's condition; however, the ranking order demonstrated that parents with a child who had CHD considered these questions to be top priority.
- Health: Most parents were concerned about a wide range of health issues, and this was ranked the highest by parents who had children with DS.
- Education: There was consensus among all parents about the importance of education, and this was particularly important in the ranking for parents of children with SB.
- Psychosocial impact: This issue was ranked to be of equal importance across all groups.

Table 2. Ten most important research questions of ConnectEpeople survey respondents with children who have Down syndrome (N=35).

Question	Really important/important, n (%)	Not sure, n (%)	Definitely not important, n (%)
How can I maximize my child's educational attainment?	34 (97)	1 (3)	0 (0)
What dietary supplements should my child be taking?	34 (97)	1 (3)	0 (0)
Does exercise enhance the immune system of children with Down syndrome?	33 (94)	1 (3)	1 (3)
Would early intervention, eg, tummy time, creeping, and crawling, enhance my child's development?	33 (94)	1 (3)	1 (3)
How many children, with the same condition as my child, go to mainstream school?	32 (91)	1 (3)	2 (6)
Where would I find specialized information such as video clips of parents feeding a baby with my child's condition?	32 (91)	0 (0)	3 (9)
Is obesity a problem with my child's condition?	32 (91)	0 (0)	3 (9)
What is the latest genetic research relating to my child's condition?	32 (91)	3 (9)	0 (0)
What is the psychosocial impact of my child's condition on my child and our family?	32 (91)	1 (3)	2 (6)
What complementary therapies are beneficial for my child?	32 (91)	2 (6)	1 (3)

Table 3. Ten most important research questions of ConnectEpeople survey respondents with children who have spina bifida (N=39).

Question	Really important/important, n (%)	Not sure, n (%)	Definitely not important, n (%)
How many children have surgery and how many survive?	37 (95)	2 (5)	0 (0)
What is the psychosocial impact of my child's condition on my child and our family?	37 (95)	2 (5)	0 (0)
What is the normal milestone development for a child with the same condition as my child?	36 (92)	2 (5)	1 (3)
If my child has to take time out of school, will their education continue?	36 (92)	3 (8)	0 (0)
How can I maximize my child's educational attainment?	36 (92)	3 (8)	0 (0)
What complementary therapies are beneficial for my child?	36 (92)	2 (5)	1 (3)
What dietary supplements should my child be taking?	35 (90)	3 (8)	1 (2)
What devices or products are the best to buy for my child at different life stages?	35 (90)	3 (8)	1 (2)
How many children, with the same condition as my child, go to mainstream school?	35 (90)	3 (8)	1 (2)
What type of operations are available for babies in the womb to reduce the effect of their condition?	34 (87)	4 (10)	1 (3)

Table 4. Ten most important research questions of ConnectEpeople survey respondents with children who have cleft lip with or without cleft palate (N=16).

Question	Really important/important, n (%)	Not sure, n (%)	Definitely not important, n (%)
What is the rate of reoccurrence of cleft lip with or without cleft palate among siblings?	16 (100)	0 (0)	0 (0)
Are there lactation consultants with expertise in supporting parents who have a child like mine?	16 (100)	0 (0)	0 (0)
What are the genetic and environmental causes of cleft lip with or without cleft palate?	16 (100)	0 (0)	0 (0)
Where would I find specialized information such as video clips of parents feeding a baby with my child's condition?	16 (100)	0 (0)	0 (0)
What is the best age for children with a cleft to have surgery?	16 (100)	0 (0)	0 (0)
What is the latest genetic research relating to my child's condition?	15 (94)	1 (6)	0 (0)
How can I maximize my child's educational attainment?	14 (88)	2 (12)	0 (0)
What complementary therapies are beneficial for my child?	14 (88)	2 (12)	0 (0)
What is the psychosocial impact of my child's condition on my child and our family?	14 (88)	0 (0)	2 (12)
What is the normal milestone development for a child with the same condition as my child?	13 (81)	2 (13)	1 (6)

Table 5. Ten most important research questions of ConnectEpeople survey respondents with children who have congenital heart defects (N=10).

Question	Really important/important, n (%)	Not sure, n (%)	Definitely not important, n (%)
If my child is diagnosed with a heart condition in the womb, are there any medications I can take to help my baby?	10 (100)	0 (0)	0 (0)
Is it okay for my child to get vaccinated?	10 (100)	0 (0)	0 (0)
Is there an increased number of hospital admissions during winter with children with heart defects?	10 (100)	0 (0)	0 (0)
Is obesity a problem with my child's condition?	10 (100)	0 (0)	0 (0)
What is the latest genetic research relating to my child's condition?	10 (100)	0 (0)	0 (0)
How can I maximize my child's educational attainment?	10 (100)	0 (0)	0 (0)
What is the psychosocial impact of my child's condition on my child and our family?	10 (100)	0 (0)	0 (0)
Can you pick up heart defects during pregnancy and reduce the damage?	9 (90)	1 (10)	0 (0)
How many children have heart surgery and how many survive?	9 (90)	1 (10)	0 (0)
What age is my child likely to live to?	9 (90)	1 (10)	0 (0)

Discussion

Principal Findings

The key finding from this social media research is that parents with children who have CAs value social media for connecting with others and to obtain information about their child's condition, future health, well-being, educational outcomes, and psychosocial issues. This study affirms growing parental preference for information and support via interactive social media and less interest in commonly perceived useful sources of support for parents, such as advice chat lines (online or telephone).

Social Media

By using secret Facebook groups, the ConnectEpeople e-forum enabled a process of engaging in a consultative dialogue online between researchers and parents living in European countries who have children with DS, CLP, SB, and CHD. Parents were able to contribute and collaborate to identify and prioritize research questions in a private, confidential, and secure online community. The Web-based survey respondents clearly identified social media as a popular and preferable method of communicating with others. Online discussion forums were important for parents to communicate and connect with others with shared life experiences (living with children who have CAs). The traditional method of connecting, namely, meeting in person, remains desirable when communicating with others; therefore, our approach of connecting online, using visual technology (WhatsApp, Skype, and other forms of video chat) to overcome distance, enabled the parents to *see* each other and connect with a person as opposed to making a connection with a text or a voice. Theoretically, this is what we value in ocularcentrism, where people need to see each other first and then they can communicate more effectively using text, etc.

Heath et al [24] identified that the choice of communication method can greatly enhance the research participant's willingness and ability to speak openly and honestly. Flexibility in the data collection approach from RAPs adopted by the ConnectEpeople study has yielded valuable and comprehensive information on parent's research priorities.

The ConnectEpeople survey was presented as a chance for respondents to voice their research wants and needs. To demonstrate that the parent's views were highly valued, respondents were given the opportunity to submit a research question of their own. Analysis of the additional research questions collected in the open question revealed themes consistent with the questions presented to and prioritized by survey respondents. This suggests that still much needs to be done globally to ensure that all parents of children with CAs have timely access to robust evidence-based information that they feel they need to meet the needs of their children and provide them with the best possible opportunities, care, and treatments.

Ocularcentrism

From a theoretical perspective, understanding how and why people value visible data is an important determinant in their preference for the use of social media. The term *ocularcentrism* has become familiar to researchers in social media and is a behavior that is built on the theory that *seeing* is believing and that we naturally favor visual contact [16]. In this research, we see parents who use "technology that is manifested in the use of social media" possibly because of its illustrative and graphical potential to optimize visual representation in person first, then in word, text, and audio [16]. The online behavior is almost second nature to the expert user who controls the software, as if they drive a car, and after a time, the behavior techniques become so normalized and intrinsic that they are almost subconscious. Current research highlights the preferences by the public for using social media [25]. Although the findings of the ConnectEpeople survey confirm this, it sheds light on the high value that parents continue to place on traditional meeting in-person communication. Therefore, given that online video chat is widely accessible, convenient, and inexpensive, it is a valuable component to facilitate meeting in person for collaborative research with geographically distant teams.

Face-to-face contact is significant in building trust and rapport and enabling team members to speak their mind. Therefore, developing opportunities for research participants in other countries to meet researchers face-to-face online will facilitate a higher level of collaboration and engagement and is in keeping with our understanding of the value we place on *seeing* not only words, text, and video but also the human person and is in keeping with our reference to ocularcentrism [16].

Research Priorities

We expected, based on our review of the literature, that by working with parents, very different views and needs would be expressed. Therefore, the decision was taken to use the JLA approach, which has demonstrated expertise in the area of identifying research priorities. The aim of this work was to give parents who have children with CAs a platform to voice their opinion on what issues are important to them and rank them in the order of importance.

Parents had a varied response to the identification of key research questions, and RAPs who reviewed the lists of questions were *not surprised* by the findings and agreed with the value and importance of the research questions generated and the rankings. Maintaining good health, maximizing educational attainment, and improving psychosocial aspects were among the top concerns across all 4 CAs. The areas of interest for respondents and the parents who developed the list of questions were concerned with a range of issues, eg, achieving childhood developmental milestones, infant feeding, complementary therapies, and exercise. There may be a lack of available information for parents in these areas that would resonate with the research by McHugh et al [26] who made a strong argument for collaborative research for children with chronic health conditions. Their work demonstrated that parents' and investigators' research needs are often incongruent and researchers do not clearly understand the issues that are important to parents. Parents and carers of children participating in research need to be involved in the process of prioritizing research questions. Morris et al [27] demonstrated the benefits of a British Academy of Childhood Disability, JLA Research PSP for children with disabilities, working with a wide range of stakeholders, including parents. Together, the multiprofessional team developed a *Top Ten List* of research priorities that led to the identification of funding opportunities from the NIHR and National Institute for Health and Care Excellence (NICE) guidelines. Therefore, the onus is placed on researchers and health professionals to address these information needs and keep abreast of parents' changing needs.

Additional Findings

Although the recruitment process was lengthy, it was ethically appropriate and was a clear demonstration of our sensitivity and human caring for parents. This was confirmed when organizations and parents commented on the value of the face-to-face recruitment approach and using the STAI as a measure of well-being. We report this finding as new knowledge about the value of a screening tool such as the STAI to facilitate ethical recruitment in sensitive research cases.

Limitations

The research limitations include the need for RAPs to be able to understand English, be able to access the survey online, have a child aged between 1 and 11 years, and be users of social media/technology for communication. In addition, 77% (62/80) of the survey respondents who fully completed the survey were educated to degree or postgraduate level. The survey was live from May 24, 2018, to October 8, 2018, over the holiday period, and the sample size was small.

Conclusions

The use of social media and online parental engagement in research within the ConnectEpeople project enabled the identification of parent's research priorities. Working online and using face-to-face apps and technology can build trust and foster the collaboration by exploiting multiple communication channels to maximize engagement and partnership between researchers and parents to produce accessible, meaningful, and usable information. The survey revealed that meeting in person continues to be highly valued alongside social media and discussion forums. The agreement in research priorities between the survey respondents and the RAPs and the wide geographical engagement suggest a high degree of commonality of the research wants and needs of parents of children with these CAs, regardless of global location.

Acknowledgments

The authors wish to thank all the parents involved in the study for providing research questions and generously sharing their experiences and expertise and their open and honest feedback and review. The authors also wish to thank Dr Paul Slater for his statistical advice. This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement number 733001.

Authors' Contributions

MS conceptualized the study. MS and JEMMcC led the social media data collection. All authors developed the survey tool. JEMMcC analyzed the data. MS and JEMMcC drafted the manuscript. All named authors contributed with improvements and critical revisions and approved the final version for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Mean state anxiety scores (95% CI), based on the State-Trait Anxiety Inventory, of screened parents according to the congenital anomaly of their child.
[PNG File , 13 KB - jmir_v21i11e15847_app1.png]

Multimedia Appendix 2
ConnectEpeople survey respondents' communication preferences with support group(s), networks, or charities (n=65).
[PNG File , 37 KB - jmir_v21i11e15847_app2.png]

References

1. European Commission. H2020 Programme Guidance Social media guide for EU funded R&I projects URL: http://ec.europa.eu/research/participants/data/ref/h2020/other/grants_manual/amga/soc-med-guide_en.pdf [accessed 2019-11-15]
2. Douglas T, Redley B, Ottmann G. The need to know: the information needs of parents of infants with an intellectual disability-a qualitative study. J Adv Nurs 2017 Nov;73(11):2600-2608. [doi: 10.1111/jan.13321] [Medline: 28440950]
3. DeHoff B, Staten L, Rodgers R, Denne S. The role of online social support in supporting and educating parents of young children with special health care needs in the United States: a scoping review. J Med Internet Res 2016 Dec 22;18(12):e333 [FREE Full text] [doi: 10.2196/jmir.6722] [Medline: 28007689]
4. Kirk S, Milnes L. An exploration of how young people and parents use online support in the context of living with cystic fibrosis. Health Expect 2016 Apr;19(2):309-321 [FREE Full text] [doi: 10.1111/hex.12352] [Medline: 25691209]
5. Pehora C, Gajaria N, Stoute M, Fracassa S, Serebale-O'Sullivan R, Matava C. Are parents getting it right? A survey of parents' internet use for children's health care information. Interact J Med Res 2015 Jun 22;4(2):e12 [FREE Full text] [doi: 10.2196/ijmr.3790] [Medline: 26099207]
6. Plantin L, Daneback K. Parenthood, information and support on the internet. A literature review of research on parents and professionals online. BMC Fam Pract 2009 May 18;10:34 [FREE Full text] [doi: 10.1186/1471-2296-10-34] [Medline: 19450251]
7. Nicholl H, Tracey C, Begley T, King C, Lynch A. Internet use by parents of children with rare conditions: findings from a study on parents' web information needs. J Med Internet Res 2017 Feb 28;19(2):e51 [FREE Full text] [doi: 10.2196/jmir.5834] [Medline: 28246072]
8. Lavan A, Reiter S, Heiman T. Educational involvement of parents of mainstreamed special needs children. Contemp School Psychol 2018:1-11 [FREE Full text] [doi: 10.1007/s40688-018-0202-1]
9. Amirav I, Vandall-Walker V, Rasiah J, Saunders L. Patient and researcher engagement in health research: a parent's perspective. Pediatrics 2017 Sep;140(3):pii: e20164127 [FREE Full text] [doi: 10.1542/peds.2016-4127] [Medline: 28851740]
10. European Lung Foundation. Patient and Public Involvement URL: https://www.europeanlung.org/assets/files/en/publications/patientinput.pdf [accessed 2019-11-15]
11. Lander J, Langhof H, Dierks M. Involving patients and the public in medical and health care research studies: an exploratory survey on participant recruiting and representativeness from the perspective of study authors. PLoS One 2019;14(1):e0204187 [FREE Full text] [doi: 10.1371/journal.pone.0204187] [Medline: 30615619]
12. INVOLVE. Public Information Pack (PIP): How to Get Involved in NHS, Public Health and Social Care Research URL: https://www.invo.org.uk/wp-content/uploads/2019/04/PIP12019.pdf [accessed 2019-11-15] [WebCite Cache ID 78aya6C6k]
13. The United Nations Children's Fund (UNICEF). 2015. Online Communities URL: https://www.unicef.org/knowledge-exchange/files/Online_Communities_production.pdf [accessed 2019-11-15]
14. Russell DJ, Sprung J, McCauley D, de Camargo OK, Buchanan F, Gulko R, et al. Knowledge exchange and discovery in the age of social media: the journey from inception to establishment of a parent-led web-based research advisory community for childhood disability. J Med Internet Res 2016 Nov 11;18(11):e293 [FREE Full text] [doi: 10.2196/jmir.5994] [Medline: 27836818]
15. Sinclair M. Midwives? Readiness to Use High Technology in the Labour Ward. Unpublished PhD Thesis, Queen?s University, Belfast (p191) 1999:191-191.
16. Sinclair M. Occularcentrism and the need to 'see' the evidence of impact. Evid Based Midwifery 2011;9(2):39-40 [FREE Full text]
17. EUROlinkCAT. 2017. What is EUROlinkCAT? URL: https://www.eurolinkcat.eu/eurolinkcat/whatiseurolinkcat [accessed 2019-11-20]
18. EUROlinkCAT. Parents Area URL: https://www.eurolinkcat.eu/forparentsarea [accessed 2019-11-15]
19. Sinclair M, Latos-Bielenska A, McCullough JE, Elliott D. ConnectEPeople: a review of the online support organizations across the EU for parents with children who have down syndrome, spina bifida, cleft lip and severe heart disease requiring surgery. Eur J Med Genet 2018;61(9):570-571 [FREE Full text] [doi: 10.1016/j.ejmg.2018.06.067]
20. Spielberger CD. State - Trait Anxiety Inventory. In: The Corsini Encyclopedia of Psychology. Hoboken, NJ: Wiley; 2010.
21. Sesti A. State Trait Anxiety Inventory (STAI) in Medication Clinical Trials. Quality of Life Newsletter. ? 2000;25:17.

22. Gorenstein C, Andrade L. Validation of a Portuguese version of the Beck Depression Inventory and the State-Trait Anxiety Inventory in Brazilian subjects. Braz J Med Biol Res 1996 Apr;29(4):453-457. [Medline: 8736107]
23. The James Lind Alliance. URL: http://www.jla.nihr.ac.uk/ [accessed 2019-05-30] [WebCite Cache ID 78kvzw3Aq]
24. Heath J, Williamson H, Williams L, Harcourt D. 'It's just more personal': using multiple methods of qualitative data collection to facilitate participation in research focusing on sensitive subjects. Appl Nurs Res 2018 Oct;43:30-35. [doi: 10.1016/j.apnr.2018.06.015] [Medline: 30220360]
25. Perrin A, Anderson M. Pew Research Center. 2019 Apr 10. Share of US Adults Using Social Media, Including Facebook, is Mostly Unchanged Since 2018 URL: https://www.pewresearch.org/fact-tank/2019/04/10/share-of-u-s-adults-using-social-media-including-facebook-is-mostly-unchanged-since-2018/ [accessed 2019-11-15]
26. McHugh C, Bailey S, Shilling V, Morris C. Meeting the information needs of families of children with chronic health conditions. Phys Occup Ther Pediatr 2013 Aug;33(3):265-270. [doi: 10.3109/01942638.2013.799628] [Medline: 23701532]
27. Morris C, Simkiss D, Busk M, Morris M, Allard A, Denness J, et al. Setting research priorities to improve the health of children and young people with neurodisability: a British Academy of Childhood Disability-James Lind Alliance Research Priority Setting Partnership. BMJ Open 2015 Jan 28;5(1):e006233 [FREE Full text] [doi: 10.1136/bmjopen-2014-006233] [Medline: 25631309]

Abbreviations

CA: congenital anomaly
CHD: congenital heart defects
CLP: cleft lip with or without cleft palate
DS: Down syndrome
e-forum: electronic forum
EUROlinkCAT: Establishing a linked European Cohort of Children with Congenital Anomalies
JLA: James Lind Alliance
NIHR: National Institute for Health Research
PPI: patient and public involvement
PSP: Priority Setting Partnership
RAP: research-aware parent
SB: spina bifida
STAI: State-Trait Anxiety Inventory

Edited by G Eysenbach; submitted 14.08.19; peer-reviewed by M Herron, G Kernohan; comments to author 20.09.19; revised version received 29.09.19; accepted 30.09.19; published 25.11.19.

Please cite as:
Sinclair M, McCullough JEM, Elliott D, Latos-Bielenska A, Braz P, Cavero-Carbonell C, Jamry-Dziurla A, João Santos A, Páramo-Rodríguez L
Exploring Research Priorities of Parents Who Have Children With Down Syndrome, Cleft Lip With or Without Cleft Palate, Congenital Heart Defects, or Spina Bifida Using ConnectEpeople: A Social Media Coproduction Research Study
J Med Internet Res 2019;21(11):e15847
URL: http://www.jmir.org/2019/11/e15847/
doi:10.2196/15847
PMID:31763986

Review

Qualitative Synthesis of Young People's Experiences With Technology-Assisted Cognitive Behavioral Therapy: Systematic Review

Darragh McCashin[1], BA, HDip, MSc; David Coyle[2], PhD; Gary O'Reilly[1], PhD

[1]School of Psychology, University College Dublin, Dublin, Ireland

[2]School of Computer Science, University College Dublin, Dublin, Ireland

Corresponding Author:
Darragh McCashin, BA, HDip, MSc
School of Psychology
University College Dublin
Belfield
Dublin, Dublin 4
Ireland
Phone: 353 1718363
Email: darragh.mccashin@ucd.ie

Abstract

Background: Cognitive behavioral therapy (CBT) for young people is increasingly being provided using technology-assisted formats. Although there is increasing evidence regarding the efficacy of such approaches, as illustrated by quantitative systematic reviews, the literature has also highlighted challenges with implementation factors, including high attrition rates and variable user engagement. Qualitative review methods can help to address the factors that impact young peoples' experience of technology-assisted cognitive behavioral therapy (tech-assisted CBT) and, thus, enable us to better understand such implementation factors. To date, no such qualitative synthesis exists.

Objective: The primary aim of this review was to systematically identify and synthesize the qualitative literature concerning the experiences of young people who have used tech-assisted CBT.

Methods: This systematic review applied Thomas and Harden's 2008 qualitative thematic synthesis approach. This involved line-by-line coding of the results sections of included studies and an inductive analysis on identified themes, followed by the generation of analytical themes through a process of iteration and interpretation of the descriptive themes. PsycINFO, ACM Digital Library, PubMed, EMBASE, and JMIR Publications databases were searched. The inclusion criteria were (1) studies involving school-aged young people over preschool age (6 years) but under the age of 18 years, (2) use of any form of tech-assisted CBT for any time period, (3) a stated focus of qualitative data to document the experiences of participants, and (4) studies published in English. The exclusion criteria were (1) interventions only provided face-to-face with no technological component, (2) only focused on the performance of the technology rather than participant experience, and (3) numerical data that sought to represent qualitative data.

Results: A total of 14 studies were included in this review. Overall, these studies represented interventions for low mood and anxiety (n=10), trauma or self-harm (n=2), and physical difficulties (n=2). Overall, 5 analytical themes emerged on young people's experiences with tech-assisted CBT: (1) helpfulness, (2) therapeutic process, (3) transferability, (4) gameplay experience, and (5) limitations. In addition, these analytical themes contained the following subthemes: positive experiences, tech-assisted CBT versus face-to-face CBT, understanding of a CBT model, process of change, skills development, application to everyday life settings, parental involvement, character relatedness, playability, negative experiences, and broad content.

Conclusions: Overall, young people's experiences with tech-assisted CBT were mostly positive. The use of gaming environments, relatable characters, concrete metaphors, and age-appropriate narratives contributed to these positive experiences. Evidence suggests that technology can help to mediate face-to-face relationships with therapists and help young people to understand the CBT model. Clear barriers also emerged, including over-reliance on reading and writing skills and dissatisfaction with overly generalized content and comparison with commercial technologies.

Trial Registration: International Prospective Register of Systematic Reviews (PROSPERO) CRD42018103388; https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42018103388

(J Med Internet Res 2019;21(11):e13540) doi:10.2196/13540

KEYWORDS

cognitive behavioral therapy; systematic review; qualitative research; children; mental health; technology; mHealth; eHealth

Introduction

Background

Cognitive behavioral therapy (CBT) is used in a range of psychological services for different populations. Given the weight of evidence underpinning CBT [1], it has been recommended as the first response to low levels of anxiety and depression [2]. However, because of resource limitations within mental health services and the notable differences to consider when providing CBT to younger populations [3], CBT has been blended with technology to provide greater accessibility. Digital CBT for depression in young people is now provisionally recommended by the National Institute for Health and Care Excellence as the first line of intervention [4]. Multidisciplinary fields, including psychology, psychiatry, health informatics, and human-computer interaction (HCI), have demonstrated the effectiveness of a range of technology-assisted cognitive behavioral therapy (tech-assisted CBT) interventions for adults and young people across different settings, primarily using quantitative approaches, including formalized trials and systematic reviews [5]. Comparatively, there have been limited systematic qualitative reviews concerning the experiences of young people using tech-assisted CBT. This review commences with a brief clarification on the operationalized definition of tech-assisted CBT, followed by a theoretical summary of CBT. Thereafter, the current gap in the literature is outlined in the context of the limitations of effectiveness research. Finally, this review addresses this gap by applying a qualitative thematic synthesis approach to addressing the experiences of young people using tech-assisted CBT in what is, to our knowledge, the first of its kind to date. This is of importance for future tech-assisted CBT design and evaluations by contextualizing how and why tech-assisted CBT interventions are effective across different settings from the perspective of the young people who have used it.

Defining Technology-Assisted Cognitive Behavioral Therapy

The introduction of first-generation forms of tech-assisted CBT, commonly referred to as computerized cognitive behavioral therapy (cCBT), was motivated by the need to address the accessibility challenges within mental health services provision. Where demand for such services exceeded the supply, cCBT was primarily envisioned to be a viable method to enhance the reach of evidence-based low-intensity interventions, in addition to being a cost-effective means of doing so [6]. Many of these interventions targeted adults and were mainly internet-based with minimal face-to-face contact [7]. However, within this period, the focus on the potential user experience brought about through providing CBT via technology or on the potentially unintended outcomes and experiences of key stakeholders was less.

This has led to a new wave of different tech-assisted CBT interventions. In addition to acknowledging the economic and accessibility advantages of these interventions, this new wave of tech-assisted CBT has identified the need to continually optimize the design of technologies, learn from user experiences, and measure implementation variables to maximize and sustain positive outcomes [8]. In addition, there is now a wider consideration of the factors that may differentiate tech-assisted CBT from traditional face-to-face CBT, including potentially novel therapeutic barriers and facilitators [9-11]. As a result, there is now a diverse range of interventions that use different technological ingredients to support the CBT process, including mobile phone and tablet apps [12], game-based software [13], interactive websites [14], virtual reality [15], and telecommunications [16]. Increasingly, these new interventions are embedded within a stepped care model [17].

Therefore, tech-assisted CBT can be defined in a variety of ways. Indeed, several descriptors for this intervention are observed throughout the literature, including internet-CBT, cCBT, CBT apps, CBT games, tele-CBT, and virtual reality CBT. For the purposes of this review, tech-assisted CBT is broadly considered as any CBT-based intervention that uses technology to facilitate, support, supplement, or replace traditional face-to-face CBT.

Theoretical Overview

The theoretical foundation to cognitive behavioral theory posits that there is a complex relationship between one's thoughts, feelings, and behaviors (TFBs) [18]. Beck [19] suggests that, for those experiencing mental or physical difficulties, automatic surface-level thinking (level 1) can often be negative, thereby perpetuating a cycle that can sustain or intensify the initial difficulty. These problematic thinking-feeling-behavior cycles can be maintained via intermediate cognitive processes (level 2), such as memory, attention, beliefs or attitudes, and interpretation. Underpinning these processes is a schematic structure, referred to as *core beliefs*, that implicitly informs one's worldview, view of others, and view of self. To identify and address each of these interdependent cognitive levels, CBT provides a structured intervention that necessitates the use of metacognition (thinking about one's thinking) to recognize problematic patterns adversely impacting the individual. As such, CBT represents a complex exercise requiring significant psychoeducational engagement in both therapeutic and real-world settings.

The complexity of the generic CBT model is further amplified when offered to young people who present different developmental needs. For CBT to be effective with young people, it needs to be appropriately tailored to their developmental stage [20]. By using child-friendly content, tech-assisted CBT interventions have the potential to make core CBT concepts more accessible to young people. Relatedly, as the generic cognitive model is underpinned by a highly

structured information-processing paradigm, it is particularly amenable to computer-assisted implementation [21]. Indeed, the integration of age-appropriate content within this process using narrative-based metaphors and accessible characters has been suggested as a developmentally appropriate way to scaffold core CBT concepts [22].

Rationale for Qualitative Synthesis Within Systematic Reviews

The *gold standard* methodological approach favored by the human and social sciences has long been the randomized controlled trial (RCT) [23]. The procedures stipulated by RCTs allow for the minimization of bias and, thus, the detection of valid and reliable intervention effects. Within behavioral sciences, it is also recognized that high-quality RCTs can have a qualitative component embedded to meaningfully complement the overall dataset [24]. The quantitative method for evaluating the overall effectiveness of an intervention is meta-analysis. This allows for a robust and cumulative appraisal of the effectiveness of an intervention across many studies [25]. Overall, these quantitative approaches significantly contribute to building and (re)evaluating evidence bases and, thus, scientific advancement and eventual impactful policy [26]. However, the qualitative component of outcome studies has often not been included in traditional systematic reviews.

More recently, there has been an advancement of similar high-quality reviewing processes for qualitative data [27]. This has occurred in response to the sometimes restrictive nature of RCTs and what can be missed by not collecting qualitative data. In addition, evidence-based and ethical policy making increasingly necessitates the inclusion of the *voice* of those at whom a policy or intervention is aimed. Indeed, in the area of mental health policy, this is reflected in the embedded role of service user perspectives across research and government institutions [28]. Similarly, in the HCI field, this is observed in participatory or user-centered iterative design approaches [29]. These inclusive methodologies are considered especially relevant for young people, who are often not included within decision-making processes.

Although the apparent objective versus subjective nature of quantitative and qualitative data has long been debated in psychology [30], there has been a notable refinement on how to situate qualitative data within high-quality frameworks of evidence [31]. It is now more conventional to see an evidential framework for an intervention to include a meta-analysis reviewing the current status, followed by an adaptive RCT that is complemented by a qualitative evaluation with participants from within the trial. Such evaluations, therefore, allow for participant experiences, positive and negative feedback, and discursiveness to be aligned with the quantitative outcomes, thereby providing a richer dataset overall.

Despite the proliferation of tech-assisted CBT meta-analyses and RCTs, there have been no qualitative syntheses of the experiences of young people who have used such interventions. Given the vast body of studies demonstrating the feasibility, acceptability, and effectiveness of various tech-assisted CBT interventions [32], this study aimed to synthesize qualitative data from young people who have used tech-assisted CBT for a range of mental and physical issues. Furthermore, this study aimed to collate the overall qualitative themes from these studies to describe the experiences of young people using tech-assisted CBT. Thereafter, the qualitative synthesis can establish analytical themes that can further inform key concepts, understandings, or hypotheses [33]. In doing so, the findings can inform the future design and evaluation of tech-assisted CBT.

Added Value of Qualitative Synthesis

Systematic reviews that use qualitative synthesis methodology are in keeping with the *moving beyond effectiveness* trajectory within evidence synthesis [34]. Moreover, it is particularly relevant for technology-based interventions by examining the contextual and implementation factors concerning how any intervention is performed and why [8]. Therefore, this study will complement the existing wealth of quantitative studies on tech-assisted CBT by providing an insight into the overall experiences of young people who have used these interventions.

Methods

A fully accessible protocol for this review was registered on PROSPERO, the international prospective register for systematic reviews [35].

Search Strategy

A broad search strategy was used across the following databases to capture research bridging psychology, HCI, and related fields: PsycINFO, ACM Digital Library, PubMed, EMBASE, and JMIR Publications (full-text search engine).

The search strategy applied the database-specific methods of using the following search terms: (cognitive OR cognitive therapy OR CBT OR cCBT OR e-therapy OR web-assisted OR tech-assisted OR game-based OR internet) AND (child OR adolescent OR youth OR young people OR young person OR kid OR teen) AND (experience OR perspective OR view OR opinion OR feedback OR interview OR thought OR focus group OR qualitative). A full breakdown of the search terminology is provided in the appendices. In addition, of the selected studies, reference sections were hand-searched for any overlooked literature. The search was conducted in March 2018 and uploaded to Rayyan [36] for storage and screening purposes.

Inclusion and Exclusion Criteria

Given the wide applicability of tech-assisted CBT, broad inclusion criteria were chosen to facilitate the capture of a wide range of data. As such, the following studies were sought: Studies that are exclusively qualitative by design; studies that have a significant qualitative component alongside quantitative measures; and formal qualitative studies of established or emerging technologies that applied, or sought to apply, tech-assisted CBT.

With respect to the population, the following criteria were applied.

Inclusion Criteria

The inclusion criteria were as follows: School-aged young people under the age of 18 years and over the age of preschool

(age 6 years); use of any form of tech-assisted CBT-based program for any period (including cCBT, Web-based/assisted CBT, CBT games, CBT mobile apps, blended cCBT, and face-to-face CBT); a stated focus on qualitative data documenting the experiences of the participants that allowed them to express their thoughts, feelings, reflections, both positive and negative feedback, and overall perspectives; and studies published in English.

Exclusion Criteria

The exclusion criteria were as follows: CBT is delivered only face to face, with no technological component; only focused on the performance of the technology rather than the participant experience; and numerical data that seek to represent qualitative data.

Screening and Data Extraction

Owing to the difficulty of locating qualitative data within major databases [37], both the titles and abstracts were screened by DM for potential eligibility. For example, RCTs embed rich qualitative evaluations but may not make this explicit in study abstracts. Following this process, full papers were assessed for eligibility wherein papers that did not utilize qualitative methods could be excluded, and any indecision was recorded via Rayyan. Using the eligibility criteria, 2 independent researchers screened 10% of the overall search and screening by DM. The resulting indecision and disagreement (2%) were discussed and clarified by the review team until consensus was reached.

Thereafter, author DM extracted all data within the *results* or *findings* section that contained direct quotations from the population. It was decided that, as many research papers apply an interpretative voice to qualitative data, this could skew the *voice* of the young person. Therefore, only direct quotes from participants in studies meeting the inclusion criteria were extracted. All quotations were extracted and entered into the qualitative software program NVivo 12 [38].

Quality Assessment

There is considerable debate surrounding the quality assessment of qualitative research, and even if this can ever be appropriately appraised [39]. There is contention about whether quality assessment should be used to exclude lower-quality studies or to offer a means of assessing the weight of different included studies, given that lower-quality studies can still generate new insights [40]. As there is no consensus regarding methods for excluding studies on the grounds of their interpreted quality [33], all identified studies were included in this review. However, the quality of all included studies was assessed using the 7 quality criteria proposed by Harden et al [41] (see Multimedia Appendix 1), whereby quality is assessed within the context of young people's views. This refers to the extent to which studies were capable of addressing the review question concerning young people's overall tech-assisted CBT experiences. All studies were critically appraised on the criteria relating to the theoretical framework, clear statements of study aims, clear context description, clear description of sample and its recruitment, clear description of methods and analysis, validity and reliability, and originality of data.

Thematic Synthesis

We have selected the thematic synthesis approach laid out by Thomas and Harden [33]. This method involves (1) a line-by-line coding of each papers' results section, (2) identification of descriptive themes, and (3) generating analytical themes that extend the initial findings. All quotations were coded line by line and the development of descriptive themes occurred by employing an inductive approach using the *Query* and *Explore* functions within *NVivo*. Following the grouping of these descriptive codes, analytical themes were developed by DM alongside the research team using an iterative process of interpretation and reinterpretation of all descriptive themes and supporting quotations. In line with the Thomas and Harden approach, a process of inference throughout these steps generated the analytical themes for the overall experiences of young people and tech-assisted CBT. The resulting analytical themes should be interpreted by the variety and richness of experiences reported rather than the quantification of quotes.

Results

Overview

The search strategy produced 3214 records. Using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses flowchart (Figure 1), 14 studies were deemed eligible for inclusion. These 14 studies comprised data that investigated the experiences of young people with tech-assisted CBT for low mood or anxiety (n=10), trauma or self-harm (n=2), and physical difficulties (n=2). All studies were published between 2013 and 2018. A significant portion of the studies originated from New Zealand (n=6), but there was also broader international coverage with studies from the United States (n=2), the United Kingdom (n=2), the Republic of Ireland (n=1), Sweden (n=1), South Africa (n=1), and Spain (n=1). Of these studies, 9 different types of tech-assisted CBT interventions were investigated within a combined sample of 289 young people. A full breakdown of the characteristics for each study is provided in Multimedia Appendix 2 [42-68] with details of the intervention type, aims and context in which it was applied, participant characteristics, methods, and key themes reported.

Overall, 5 analytical themes describing young people's experiences of tech-assisted CBT arose, with each of these containing several descriptive themes. To illustrate this synthesis, a table of selected supporting quotations from across the studies is provided (Table 1) in addition to a thematic summary diagram (see Multimedia Appendix 3).

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses flowchart.

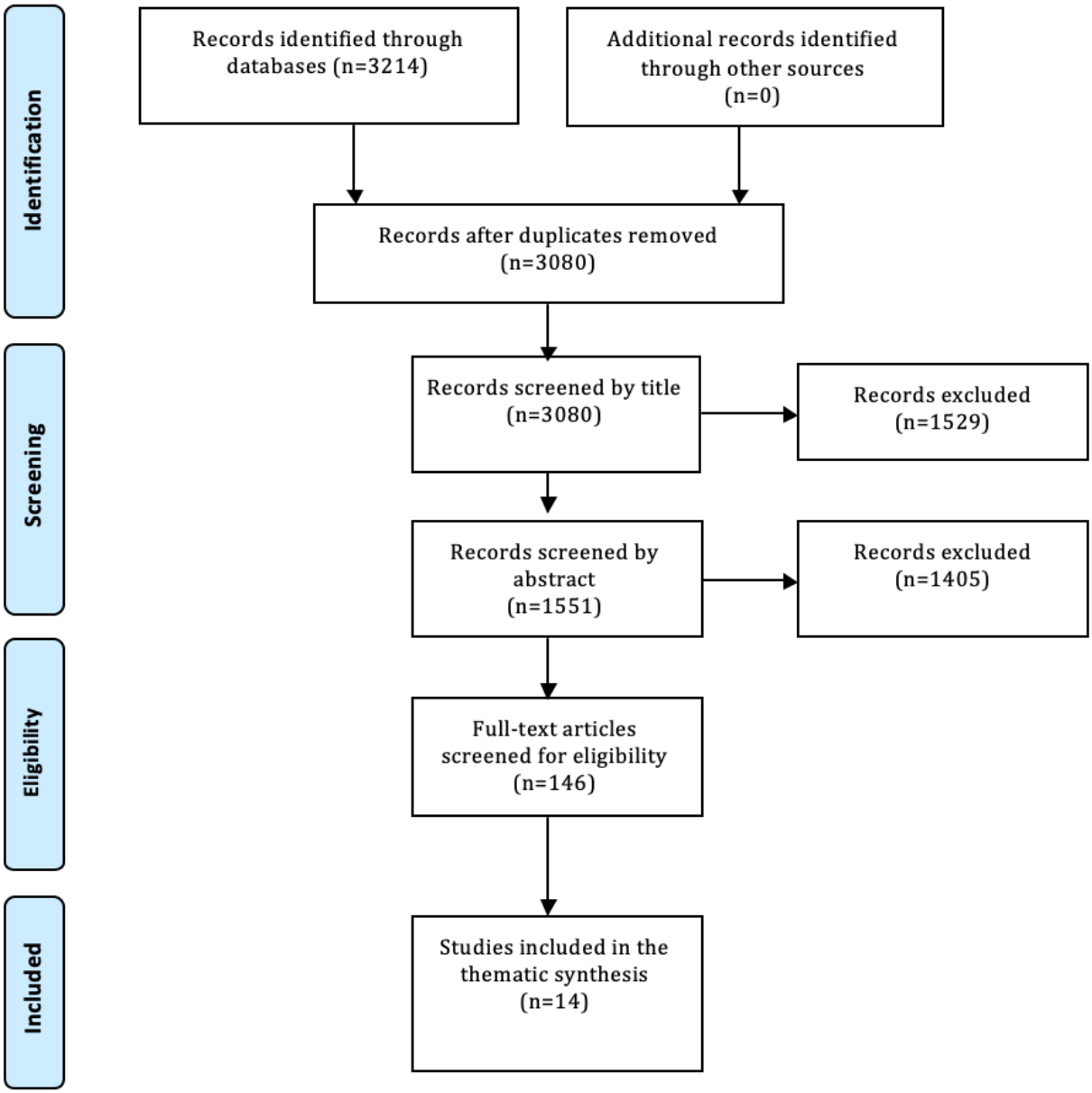


Table 1. Supporting quotation for 5 analytical themes and their descriptive subthemes.

Analytical themes	Quotations
1) The helpfulness of tech-assisted cognitive behavioral therapy (CBT) for young people	
Positive experiences	"It was really good and helpful. Like when I needed to calm down using the technique, it actually helped a lot. Yes, so overall it was pretty good, the breathing in out one." [54]
Tech-assisted CBT versus face-to-face CBT	"It's just cool...it's a different way...because you know, you go to a counselor and stuff and they have all these different ways of doing things but like, nobody's ever really thought of a computer game or something. It's usually like 'tell me how you're feeling,' or 'write it down' and stuff, but not 'play it'." [49]
2) The therapeutic process within tech-assisted CBT for young people	
Understanding CBT model	"I really liked the diary 'cos...I dunno, 'cos especially when I'm feeling down I'm like 'Oh I'm always so sad' or 'what is the point' but actually if I look back to the diary I can see that there were days when I was happy." [66]
Process of change	"Oh yeah, I used to get some of those Gnats [gloomy negative automatic thoughts]—but now I am all positive. I am thinking positive, thinking of all the good stuff that has happened to me. I don't even think about suicide or self-harm now." [13]
Skills development	"If I take myself out of the situation and just go sit somewhere else just by myself and focus on breathing, that helps a lot." [68]
3) The transferability of tech-assisted CBT for young people	
Application to everyday life settings	"I have used BlueIce every day. I used the mood checker every day and found it quite easy to use." [66]
Parental involvement	"Well, like during the holidays me and my Mum are going to do it together because she is going through a lot of stress as well. We would both sit down and do it and sort it out. I told Mum about the thing [SPARX] and she said that she would like to try it. I am hoping it will actually help her out and take less off her shoulders." [54]
4) The tech-assisted CBT gameplay experience for young people	
Character relatedness	"The thing I kind of liked about it [CCAL] was that he [referring to Charlie] was kind of going through some anxiety too, so you could see it from his point of view." [58]
Playability	"I play heaps of games, there is lots of actions, compared with other games SPARX is not great." [13]
5) The limitations of tech-assisted CBT for young people	
Negative experiences	"I might just sort of feel a bit wilful, sort of, somewhat want to stop myself from self-harming but sometimes I just want to self-harm and that's the end of it really." [66]
Broad content	"It was hard because it was catered towards a wider audience and because of that I wasn't as confident that it would help my problem." [68]

Theme 1: Helpfulness of Technology-Assisted Cognitive Behavioral Therapy for Young People

This theme outlined the varied ways in which young people experienced tech-assisted CBT to be helpful. Young people reported a range of positive experiences in using the technology for mental or physical health difficulties. This analytical theme contains the descriptive subthemes of positive experiences and tech-assisted CBT versus face-to-face CBT.

Positive Experiences

There were a large number of positive experiences reported by young people in several studies related to why they found interventions helpful (study numbers 1, 4, 6, 7, 9, and 10). These included a sense of enjoyment and fun during tech-assisted CBT sessions, meaningfulness, feeling motivated, feeling safe, hopefulness, and flexibility.

Technology-Assisted Cognitive Behavioral Therapy Versus Face-to-Face Cognitive Behavioral Therapy

It was evident that some young people were engaged with tech-assisted CBT because of their expressed preference for it compared with traditional face-to-face interventions. The ability to use tech-assisted CBT (or adjuncts to tech-assisted CBT, such as apps) encouraged help-seeking behaviors among young people (5 and 12). Moreover, young people felt that technology was easier to engage with than speaking with adults; this was because of the feeling of less stigmatization, the ability to control the pace of the game, and disliking talking (1, 3, 4, 7, and 10).

Theme 2: Therapeutic Process Within Technology-Assisted Cognitive Behavioral Therapy for Young People

This next theme provided an insight into the ways in which the therapeutic process was observed in young people's reported experiences. The underlying descriptive themes were understanding the CBT model, process of change, and skills development.

Understanding the Cognitive Behavioral Therapy Model

Numerous studies illustrated that young people understood key CBT concepts (1, 6, 7, and 12). Young people spoke about the role of thinking about one's thinking and recognizing negative thoughts and feelings without judging oneself.

Process of Change

Related to the above theme, young people were evidently applying CBT concepts to enact change as well. Several studies highlighted how young people's experiences of applying positive coping mechanisms and noticing negative automatic thoughts before they manifested as negative behaviors led to positive change in their home, school, or family life (3, 6, 12, and 14). Another study using tech-assisted CBT for trauma indicated that young people were able to utilize tech-assisted CBT technology to confront difficult memories to elicit the therapeutic process of understanding and overcoming them (11).

Skills Development

Participants across the studies indicated that many of the skills taught within tech-assisted CBT were directly therapeutic to them. Young people spoke about how focusing on breathing or progressive muscular relaxation exercises were all very helpful and made them feel calmer (6, 8, and 14).

Theme 3: Transferability of Technology-Assisted Cognitive Behavioral Therapy for Young People

The recognition from young people about the potential transferability of tech-assisted CBT in both their lives and the lives of others was a further analytical theme in the data. Young people communicated a variety of ways in which tech-assisted CBT could be transferred, as explored in 2 descriptive themes: application to everyday life settings and parental involvement.

Application to Everyday Life Settings

Young people were able to reflect on their own experiences with tech-assisted CBT. They also believed tech-assisted CBT could be applied in everyday life settings (1, 3, 6, and 12). Nonetheless, young people were mindful of some caveats to this. Young people mentioned that having components of tech-assisted CBT to share with others would be beneficial, but that some aspects of the content they used within tech-assisted CBT may not be applicable in some settings, such as school or family settings (unless privacy and discreteness were ensured). Indeed, young people acknowledged that tech-assisted CBT is applicable for many young people experiencing mental or physical health difficulties, but not all young people.

Parental Involvement

A small number of young people noticed the potential concurrent role that tech-assisted CBT could play for their parents, either as a support for their own progress or as an intervention specifically for parents (5, 10, and 12).

Theme 4: Technology-Assisted Cognitive Behavioral Therapy Gameplay Experience for Young People

The fourth theme in the synthesis is related to the features implicit in the gameplay or user experience of the respective technologies in this review. The 2 underpinning descriptive themes were character relatedness and playability.

Character Relatedness

Young people were mostly appreciative of the characters embedded within many of the tech-assisted CBT interventions (1, 3, 4, 5, 6, and 9). Many felt that the characters were relatable and that they allowed the communication of CBT content to, therefore, be easier. Notably, in studies concerning minority groups (1 and 3-6), this was especially relevant to participants who were pleased with the inclusion of Maori identities in the game. In addition, this descriptive theme extended to the personalization of young people's own role within the game. The autonomy afforded to young people in selecting their interactions with characters or personalizing avatars was considered an important gameplay feature.

Playability

Though only a minor theme, there were some mixed experiences from young people in relation to the actual playability of some tech-assisted CBT interventions (1, 3, 4, 6, 7, and 12). Some young people expressed the view that tech-assisted CBT was lacking in its gaming experience, especially when compared with commercial games. However, other young people felt that playing tech-assisted CBT games was easy, familiar, and intuitive.

Theme 5: Limitations of Technology-Assisted Cognitive Behavioral Therapy for Young People

This final theme gave an insight into the limitations of tech-assisted CBT that young people experienced. Despite many positive findings in the preceding themes, young people were also aware of the drawbacks of tech-assisted CBT. Within this, 2 descriptive themes were recorded: negative experiences and broad content.

Negative Experiences

There were a number of studies with a notably small number of young people that outlined some negative experiences with tech-assisted CBT (1, 4, 7, 8, and 12). Some young people stated that there remains a stigma associated with using tech-assisted CBT and that it is challenging to discuss psychological or physical problems with an adult, irrespective of the presence of tech-assisted CBT. Other negative experiences that were mentioned included feeling worse on occasions, not experiencing change, feeling frustrated when returning to face-to-face therapy, and feeling distracted.

Broad Content

Finally, a negligible descriptive theme arose in relation to the nature of the content in tech-assisted CBT (5, 7, and 14). Some young people identified the broad content of the psychoeducational materials to be problematic and sometimes confusing. Similarly, too much reading and writing content was also seen as a limitation of tech-assisted CBT.

Discussion

Why Do These Themes Matter?

The aim of this study was to synthesize the qualitative literature regarding the experiences of young people who have used tech-assisted CBT for a range of mental or physical difficulties. Using a qualitative synthesis approach, 5 analytical themes and 11 descriptive themes emerged from the included studies containing a diverse sample of young people. Overall, these

themes can be understood to be conceptually interrelated and are of interest for several reasons.

Taken together, themes 1 through 4 can be interpreted to broadly support the theoretical foundations of CBT, as suggested by the supporting quotation (Table 1). Young people's reported experiences of using exercises connecting TFBs were associated with themes of helpfulness and the therapeutic process. Many young people were able to think about the connections between their surface-level TFBs ("Oh yeah, I used to get some of those gnats") and engage with deeper cognitive processes including memory and attention ("...actually if I look back to the diary I can see that there were days when I was happy"), alongside some suggestions of developing more beneficial core beliefs ("but now I am all positive"). Furthermore, as effective CBT requires appropriate tailoring for young people [20], this review provides evidence that tech-assisted CBT is one effectual approach of doing so. Many of the interventions that were reviewed had applied unique ways of using technology to connect with young people, including the use of personalized avatars, child-friendly narratives, and computer game–based designs, with some young people specifying these factors as relatable ("he [referring to character] was kind of going through some anxiety too").

The first theme—the helpfulness of tech-assisted CBT for young people—corroborated the dominant trend evident in the quantitative literature, that is, young people who completed tech-assisted CBT interventions were significantly more likely to report positive outcomes versus control groups [5]. This review has provided the context to these positive trends by synthesizing what young people are saying about tech-assisted CBT, thereby offering an insight into why tech-assisted CBT is proving effective. It should be noted that young people reported the following reasons as to why tech-assisted CBT was helpful for them: they described a range of positive experiences within tech-assisted CBT and appeared to favor it over traditional face-to-face interventions. Moreover, the second analytic theme demonstrated that a range of young people in different settings experienced a therapeutic process that facilitated positive change. This is in keeping with the intended theory-to-practice model applied by core CBT approaches. Indeed, that young people highlighted the unique characteristics of tech-assisted CBT over face-to-face interventions underscored the value of offering tech-assisted CBT to appeal to young people. This is noteworthy because of the many concerns about the increasing role of technology in CBT interventions and its effect on the attitudes, acceptability, and uptake from different stakeholders over time [69,70]. This synthesis broadly validates the incorporation of technology into mental health interventions from the perspective of young people's reported experiences.

A further theme that was of interest was young people's perceived transferability of tech-assisted CBT in not only their lives but also the lives of others. Related to the therapeutic process discussed above, young people were able to recognize that their experiences in completing tech-assisted CBT were directly applicable to their different roles in school and with family and friends, and they were able to evidence these with examples. By extension, although it was a minor theme, it was also felt that parental involvement could be explored further within tech-assisted CBT interventions. This is of relevance given the potential relationship between parental anxiety and child anxiety [71], but it also suggests that there could be future adaptations of tech-assisted CBT that could further engage parents or other stakeholders. It should be noted that young people provided both supportive and unsupportive feedback regarding the involvement of parents. Similarly, young people provided several caveats to the transferability of tech-assisted CBT (eg, only some of their tech-assisted CBT content would be appropriate, and there were concerns about privacy). These considerations may be suggestive of potential subgroups of young people that might benefit from different adaptations of tech-assisted CBT that could support their individualized needs and preferences to maximize the transferability of tech-assisted CBT to key areas in their lives. As concerns remain about the long-term effectiveness of tech-assisted CBT beyond their evaluative timeframes [72], these suggestions could be especially important given that tech-assisted CBT can only maintain its effectiveness if core CBT principles are transferred to everyday life. For example, relapse-prevention strategies within tech-assisted CBT necessitate the effective transfer of CBT psychoeducation to the respective areas of young people's lives. Therefore, any strategies that can support this transfer should be explored. Nonetheless, it was apparent that young people in this review were widely transferring lessons from their tech-assisted CBT experiences into different areas of their lives.

In the context of traditional mental health interventions, tech-assisted CBT is unique given its technological mode of delivery, often guided in person by appropriately trained mental health professionals. The fourth theme—the tech-assisted CBT gameplay experience—was particularly insightful regarding the benefits of using gaming technology to support interventions. There was widespread appeal of tech-assisted CBT characters for young people, which was found to be a key factor in fostering a sense of relatedness and autonomy. This was particularly noticeable in the smart, positive, active, realistic, X-factor thoughts (SPARX) adaptations for different minority groups, where character relatedness was a recurring subtheme discussed within young people's positive experiences. This is an important dynamic because of the influential role of social learning for young people [73] and reinforces the need for continued development of personalization features for tech-assisted CBT characters that span youth cultures and identities. As a result, the experiences young people described were demonstrative of enhanced engagement with CBT content, partly because of the gameplay mechanism of interaction. However, this must be considered alongside some mixed feedback regarding the playability features of tech-assisted CBT where young people observed a contrast with commercial games. This is a worthwhile comparison for tech-assisted CBT facilitators as these contrasts could develop into intervention resistance or disengagement. This also suggests that managing young people's expectations for tech-assisted CBT is important in optimizing their subsequent engagement; however, no standardized approaches for doing so were evident in this review.

The final theme emphasized some of the limitations of tech-assisted CBT from the perspective of young people. Although this was also a minor theme, young people commented on the sometimes broad content of tech-assisted CBT in addition to a small number of negative experiences—a finding echoed in the research on adult experiences [9]. The precise reasons as to why these themes arose were unclear, but it is advisable that stakeholders factor in the persistent role of stigmatization and intervention nonresponse for a minority of young people.

Limitations

Compared with the quantitative literature, there was a significant lack of qualitative studies on this topic. Consequently, only 14 studies were included in this review. Although this proved sufficient to extract a threshold of data, the combined dataset was still relatively small in comparison with meta-analyses. In addition, most of the studies involved young people who had consented to research and who had completed their intervention. This means that the data did not include young people who may have dropped out or were incapable of research participation because of psychological or physical difficulties. Although only 1 analytical theme reflected negative experiences, it is possible that these attrition challenges reduced the ability of this review to identify the negative aspects of tech-assisted CBT or new findings related to tech-assisted CBT experiences for young people. The reasons for such attrition or nonparticipation are especially important in the critical appraisal of the uptake of tech-assisted CBT but were beyond the scope of this review. Had there been qualitative studies analyzing the experiences of those who dropped out or disliked tech-assisted CBT, a richer thematic structure regarding negative experiences or unintended consequences would have emerged.

It should also be noted that a large portion of the included studies come from 1 intervention (SPARX) that has a significant evidence base demonstrating its effectiveness. As such, the overall synthesis may be disproportionately attributable to the characteristics of this intervention. Moreover, some of these studies pertained to the design or prototyping of the intervention [43,49,51], and therefore, they may be liable to desirability biases that do not reflect the experiences of real-world adherence. This disproportionate coverage also reiterates the lack of qualitative studies for a variety of tech-assisted CBT interventions across the extant literature.

A further limitation is that only direct quotes from young people were included in the analysis and not the full results sections of papers. Although this decision was made to ensure that only the voice of young people was analyzed, it may have limited the potential insights and reliability of the overall synthesis.

In addition, as this review purposively used a broad definition of tech-assisted CBT, it is not known if the reported themes would emerge with specific variants of CBT—for example, are game-based designs of tech-assisted CBT experienced significantly differently to Web-based or app-based versions?

Future Research and Recommendations

Future research would benefit from further well-designed qualitative studies that focus exclusively on the experiences of young people. Specifically, studies should endeavor to also include the experiences of young people who find tech-assisted CBT to be unhelpful or those who drop out. This would be particularly useful in ascertaining what intervention adaptations could be made or *built in* to the technology to mitigate against attrition risks. The continued debate about the merits of qualitative research could be addressed by the implementation of enhanced quality assessment for evidence syntheses. In addition, the use of standardized questions when evaluating tech-assisted CBT could enhance the comparability of reviews.

As the tech-assisted CBT experience could be both positively or negatively affected by the absence of common factors that typify face-to-face CBT, this review has provided an insight into, and support for, the suggestion that computerized therapies, such as tech-assisted CBT, have their own unique common factors [9]. In particular, the subthemes discussed in this review illustrated novel factors brought about because of the technological component of CBT (namely, *tech-assisted CBT vs face-to-face CBT*, *character relatedness*, and *playability*). Therefore, in addition to appropriately applying the underlying CBT theory, future research and design can improve positive outcomes for young people by focusing on these novel factors. In practice, this means that designers of tech-assisted CBT should consider how to best enhance the personalization process for young people, the sophistication of child-friendly and psychologically appropriate narratives or avatars, and maximizing pathways for young people to transfer their learning to everyday life. For tech-assisted CBT practitioners, future research needs to further develop our qualitative understanding of these novel factors and how they relate to the therapeutic process, successful implementation of tech-assisted CBT, and the early identification of potential intervention nonresponders.

Acknowledgments

This research was funded by Technology Enabled Mental Health for Young People (TEAM). TEAM has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement number 722561. Special thanks to TEAM colleagues Bettina Moltrecht and Hidde van der Meulen for their assistance with data screening.

Conflicts of Interest

Both DC and GOR are founders of the nonprofit company Handaxe [56] that created *Pesky gNATs and Mindful gNATs*, which were included in this review [55,56]

Multimedia Appendix 1
Critical appraisal of included studies using criteria proposed by Harden et al.
[PDF File (Adobe PDF File), 25 KB - jmir_v21i11e13540_app1.pdf]

Multimedia Appendix 2
Characteristics of included studies.
[PDF File (Adobe PDF File), 91 KB - jmir_v21i11e13540_app2.pdf]

Multimedia Appendix 3
Thematic summary diagram.
[PNG File , 254 KB - jmir_v21i11e13540_app3.png]

References

1. Herbert JD, Forman EM. Acceptance and Mindfulness in Cognitive Behavior Therapy: Understanding and Applying the New Therapies. Hoboken, NJ: John Wiley & Sons; 2011.
2. National Collaborating Centre for Mental Health. Depression in Children and Young People: Identification and Management in Primary, Community and Secondary Care. Leicester, England: British Psychological Society and RCPsych Publications; 2005.
3. Grave J, Blissett J. Is cognitive behavior therapy developmentally appropriate for young children? A critical review of the evidence. Clin Psychol Rev 2004 Aug;24(4):399-420. [doi: 10.1016/j.cpr.2004.03.002] [Medline: 15245828]
4. Wise J. Depression in children: offer digital CBT as first line treatment, says NICE. Br Med J 2019 Jan 23;364:l364. [doi: 10.1136/bmj.l364] [Medline: 30674466]
5. Pennant ME, Loucas CE, Whittington C, Creswell C, Fonagy P, Fuggle P, Expert Advisory Group. Computerised therapies for anxiety and depression in children and young people: a systematic review and meta-analysis. Behav Res Ther 2015 Apr;67:1-18. [doi: 10.1016/j.brat.2015.01.009] [Medline: 25727678]
6. Gerhards SA, de Graaf LE, Jacobs LE, Severens JL, Huibers MJ, Arntz A, et al. Economic evaluation of online computerised cognitive-behavioural therapy without support for depression in primary care: randomised trial. Br J Psychiatry 2010 Apr;196(4):310-318. [doi: 10.1192/bjp.bp.109.065748] [Medline: 20357309]
7. Cavanagh K, Shapiro DA. Computer treatment for common mental health problems. J Clin Psychol 2004 Mar;60(3):239-251. [doi: 10.1002/jclp.10261] [Medline: 14981789]
8. Mohr DC, Riper H, Schueller SM. A solution-focused research approach to achieve an implementable revolution in digital mental health. JAMA Psychiatry 2018 Feb 1;75(2):113-114. [doi: 10.1001/jamapsychiatry.2017.3838] [Medline: 29238805]
9. Knowles SE, Toms G, Sanders C, Bee P, Lovell K, Rennick-Egglestone S, et al. Qualitative meta-synthesis of user experience of computerised therapy for depression and anxiety. PLoS One 2014;9(1):e84323 [FREE Full text] [doi: 10.1371/journal.pone.0084323] [Medline: 24465404]
10. Henson P, Wisniewski H, Hollis C, Keshavan M, Torous J. Digital mental health apps and the therapeutic alliance: initial review. BJPsych Open 2019 Jan;5(1):e15 [FREE Full text] [doi: 10.1192/bjo.2018.86] [Medline: 30762511]
11. Sucala M, Schnur JB, Constantino MJ, Miller SJ, Brackman EH, Montgomery GH. The therapeutic relationship in e-therapy for mental health: a systematic review. J Med Internet Res 2012 Aug 2;14(4):e110 [FREE Full text] [doi: 10.2196/jmir.2084] [Medline: 22858538]
12. Torous J, Powell AC. Current research and trends in the use of smartphone applications for mood disorders. Internet Interv 2015 May;2(2):169-173. [doi: 10.1016/j.invent.2015.03.002]
13. Fleming TM, Bavin L, Stasiak K, Hermansson-Webb E, Merry SN, Cheek C, et al. Serious games and gamification for mental health: current status and promising directions. Front Psychiatry 2016;7:215 [FREE Full text] [doi: 10.3389/fpsyt.2016.00215] [Medline: 28119636]
14. Spek V, Cuijpers P, Nyklícek I, Riper H, Keyzer J, Pop V. Internet-based cognitive behaviour therapy for symptoms of depression and anxiety: a meta-analysis. Psychol Med 2007 Mar;37(3):319-328. [doi: 10.1017/S0033291706008944] [Medline: 17112400]
15. Scozzari S, Gamberini L. Virtual reality as a tool for cognitive behavioral therapy: a review. In: Brahnam S, Jain LC, editors. Advanced Computational Intelligence Paradigms in Healthcare 6. Virtual Reality in Psychotherapy, Rehabilitation, and Assessment. Berlin: Springer; 2011:63-108.
16. Lovell K, Cox D, Haddock G, Jones C, Raines D, Garvey R, et al. Telephone administered cognitive behaviour therapy for treatment of obsessive compulsive disorder: randomised controlled non-inferiority trial. Br Med J 2006 Oct 28;333(7574):883 [FREE Full text] [doi: 10.1136/bmj.38940.355602.80] [Medline: 16935946]
17. Green KE, Iverson KM. Computerized cognitive-behavioral therapy in a stepped care model of treatment. Prof Psychol: Res Pract 2009;40(1):96-103 [FREE Full text] [doi: 10.1037/a0012847]
18. Beck AT. Thinking and depression. I. Idiosyncratic content and cognitive distortions. Arch Gen Psychiatry 1963 Oct;9:324-333. [doi: 10.1001/archpsyc.1963.01720160014002] [Medline: 14045261]

19. Beck AT, Dozois DJ. Cognitive therapy: current status and future directions. Annu Rev Med 2011;62:397-409. [doi: 10.1146/annurev-med-052209-100032] [Medline: 20690827]
20. Carr A. What Works with Children, Adolescents, and Adults?: A Review of Research on the Effectiveness of Psychotherapy. London: Routledge; 2008.
21. Coyle D, McGlade N, Doherty G, O'Reilly G. Exploratory Evaluations of a Computer Game Supporting Cognitive Behavioural Therapy for Adolescents. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2011 Presented at: CHI'11; May 7-11, 2011; Vancouver, BA, Canada p. 2937-2946. [doi: 10.1145/1978942.1979378]
22. O'Reilly G, Pesky G. Pesky gNATs! Using computer games and smartphone apps to teach complex cognitive behavioural therapy and mindfulness concepts to children with mental health difficulties. In: Harnish RJ, Bridges KR, Sattler DN, Signorella ML, Munson M, editors. The Use of Technology in Teaching and Learning. Washington, DC: Society for the Teaching of Psychology; 2018.
23. Torgerson DJ, Torgerson CJ. Designing Randomised Trials in Health, Education and the Social Sciences: An Introduction. London: Palgrave Macmillan; 2008.
24. O'Cathain A, Thomas KJ, Drabble SJ, Rudolph A, Hewison J. What can qualitative research do for randomised controlled trials? A systematic mapping review. BMJ Open 2013 Jun 20;3(6):e002889 [FREE Full text] [doi: 10.1136/bmjopen-2013-002889] [Medline: 23794542]
25. Borenstein M, Hedges LV, Higgins JP, Rothstein HR. Introduction to Meta-Analysis. New Jersey: John Wiley & Sons; 2011.
26. Cartwright N. Evidence-Based Policy: A Practical Guide to Doing It Better. Oxford: Oxford University Press; 2012.
27. Noyes J, Booth A, Flemming K, Garside R, Harden A, Lewin S, et al. Cochrane qualitative and implementation methods group guidance series-paper 3: methods for assessing methodological limitations, data extraction and synthesis, and confidence in synthesized qualitative findings. J Clin Epidemiol 2018 May;97:49-58. [doi: 10.1016/j.jclinepi.2017.06.020] [Medline: 29247700]
28. Thornicroft G, Tansella M. Growing recognition of the importance of service user involvement in mental health service planning and evaluation. Epidemiol Psichiatr Soc 2005;14(1):1-3. [doi: 10.1017/s1121189x00001858] [Medline: 15792287]
29. Terp M, Laursen BS, Jørgensen R, Mainz J, Bjørnes CD. A room for design: through participatory design young adults with schizophrenia become strong collaborators. Int J Ment Health Nurs 2016 Dec;25(6):496-506. [doi: 10.1111/inm.12231] [Medline: 27293176]
30. Cromby J. Feeling the way: qualitative clinical research and the affective turn. Qual Res Psychol 2011 Dec 14;9(1):88-98. [doi: 10.1080/14780887.2012.630831]
31. Hannes K, Macaitis K. A move to more systematic and transparent approaches in qualitative evidence synthesis: update on a review of published papers. Qual Res 2012 May 11;12(4):402-442. [doi: 10.1177/1468794111432992]
32. Ebert DD, Zarski AC, Christensen H, Stikkelbroek Y, Cuijpers P, Berking M, et al. Internet and computer-based cognitive behavioral therapy for anxiety and depression in youth: a meta-analysis of randomized controlled outcome trials. PLoS One 2015;10(3):e0119895 [FREE Full text] [doi: 10.1371/journal.pone.0119895] [Medline: 25786025]
33. Thomas J, Harden A. Methods for the thematic synthesis of qualitative research in systematic reviews. BMC Med Res Methodol 2008 Jul 10;8:45 [FREE Full text] [doi: 10.1186/1471-2288-8-45] [Medline: 18616818]
34. Popay J, editor. Moving Beyond Effectiveness In Evidence Synthesis: Methodological Issues in the Synthesis of Diverse Sources of Evidence. London: National Institute for Health and Clinical Excellence; 2006.
35. McCashin D, O'Reilly G, Coyle D. PROSPERO - University of York. 2018. A Qualitative Synthesis of Children's Experiences With Technology-assisted Cognitive Behavioural Therapy (CBT). URL: https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=103388 [accessed 2018-07-25]
36. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev 2016 Dec 5;5(1):210 [FREE Full text] [doi: 10.1186/s13643-016-0384-4] [Medline: 27919275]
37. Finfgeld-Connett D, Johnson ED. Literature search strategies for conducting knowledge-building and theory-generating qualitative systematic reviews. J Adv Nurs 2013 Jan;69(1):194-204 [FREE Full text] [doi: 10.1111/j.1365-2648.2012.06037.x] [Medline: 22591030]
38. Richards L. Using Nvivo In Qualitative Research. Thousand Oaks, CA: Sage Publications; 1999.
39. Thomas J, Harden A, Oakley A, Oliver S, Sutcliffe K, Rees R, et al. Integrating qualitative research with trials in systematic reviews. Br Med J 2004 Apr 24;328(7446):1010-1012 [FREE Full text] [doi: 10.1136/bmj.328.7446.1010] [Medline: 15105329]
40. Shuster JJ. Cochrane collaborations. In: Higgins JP, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0. Chichester: John Wiley & Sons; 2011:126-130.
41. Harden A, Garcia J, Oliver S, Rees R, Shepherd J, Brunton G, et al. Applying systematic review methods to studies of people's views: an example from public health research. J Epidemiol Community Health 2004 Sep;58(9):794-800 [FREE Full text] [doi: 10.1136/jech.2003.014829] [Medline: 15310807]
42. Lucassen MF, Hatcher A, Fleming TM, Stasiak K, Shepherd MJ, Merry SN. A qualitative study of sexual minority young people's experiences of computerised therapy for depression. Australas Psychiatry 2015 Jun;23(3):268-273. [doi: 10.1177/1039856215579542] [Medline: 25881962]

43. Lucassen MF, Hatcher S, Stasiak K, Fleming T, Shepherd M, Merry SN. The views of lesbian, gay and bisexual youth regarding computerised self-help for depression: an exploratory study. Advances in Mental Health 2014 Dec 17;12(1):22-33. [doi: 10.5172/jamh.2013.12.1.22]
44. Lucassen MF, Merry SN, Hatcher S, Frampton CM. Rainbow SPARX: a novel approach to addressing depression in sexual minority youth. Cogn Behav Pract 2015 May;22(2):203-216. [doi: 10.1016/j.cbpra.2013.12.008]
45. Merry SN, Stasiak K, Shepherd M, Frampton C, Fleming T, Lucassen MF. The effectiveness of SPARX, a computerised self help intervention for adolescents seeking help for depression: randomised controlled non-inferiority trial. Br Med J 2012 Apr 18;344:e2598 [FREE Full text] [doi: 10.1136/bmj.e2598] [Medline: 22517917]
46. Thomas DR. A general inductive approach for analyzing qualitative evaluation data. Am J Eval 2016 Jun 30;27(2):237-246. [doi: 10.1177/1098214005283748]
47. Poznanski EO, Grossman JA, Buchsbaum Y, Banegas M, Freeman L, Gibbons R. Preliminary studies of the reliability and validity of the children's depression rating scale. J Am Acad Child Psychiatry 1984 Mar;23(2):191-197. [doi: 10.1097/00004583-198403000-00011] [Medline: 6715741]
48. Fleming T, Lucassen M, Stasiak K, Shepherd M, Merry S. The impact and utility of computerised therapy for educationally alienated teenagers: the views of adolescents who participated in an alternative education-based trial. Clin Psychol 2015 Jan 5;20(2):94-102. [doi: 10.1111/cp.12052]
49. Cheek C, Bridgman H, Fleming T, Cummings E, Ellis L, Lucassen MF, et al. Views of young people in rural Australia on SPARX, a fantasy world developed for New Zealand youth with depression. JMIR Serious Games 2014 Feb 18;2(1):e3 [FREE Full text] [doi: 10.2196/games.3183] [Medline: 25659116]
50. Boyatzis RE. Transforming Qualitative Information: Thematic Analysis and Code Development. Newbury Park, California: Sage Publications; 1998.
51. Shepherd M, Fleming T, Lucassen M, Stasiak K, Lambie I, Merry SN. The design and relevance of a computerized gamified depression therapy program for indigenous Māori adolescents. JMIR Serious Games 2015 Mar 3;3(1):e1 [FREE Full text] [doi: 10.2196/games.3804] [Medline: 25736225]
52. Gilgun JF. Qualitative research and family psychology. J Fam Psychol 2005 Mar;19(1):40-50. [doi: 10.1037/0893-3200.19.1.40] [Medline: 15796651]
53. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
54. Shepherd M, Merry S, Lambie I, Thompson A. Indigenous adolescents' perception of an eMental health program (SPARX): exploratory qualitative assessment. JMIR Serious Games 2018 Jul 5;6(3):e13 [FREE Full text] [doi: 10.2196/games.8752] [Medline: 29980495]
55. Tunney C, Cooney P, Coyle D, O'Reilly G. Comparing young people's experience of technology-delivered v face-to-face mindfulness and relaxation: two-armed qualitative focus group study. Br J Psychiatry 2017 Apr;210(4):284-289. [doi: 10.1192/bjp.bp.115.172783] [Medline: 27765773]
56. O'Reilly G, Coyle D. 2015a. Pesky gNATs: A Cognitive Behaviour Therapy for Children With Anxiety and Low Mood. Bristol: Handaxe CIC. URL: http://www.io.nihr.ac.uk/wp-content/uploads/migrated/2626.6251b4c8.PeskygNATsCBTgameFINAL.pdf [WebCite Cache ID 72fvWgHiT]
57. Chapman R, Loades M, O'Reilly G, Coyle D, Patterson M, Salkovskis P. 'Pesky gNATs': investigating the feasibility of a novel computerized CBT intervention for adolescents with anxiety and/or depression in a tier 3 CAMHS setting. Cogn Behav Ther 2016 Dec 1;9:1-22. [doi: 10.1017/S1754470X16000222]
58. Salloum A, Crawford EA, Lewin AB, Storch EA. Consumers' and providers' perceptions of utilizing a computer-assisted cognitive behavioral therapy for childhood anxiety. Behav Cogn Psychother 2015 Jan;43(1):31-41. [doi: 10.1017/S1352465813000647] [Medline: 23886438]
59. Khanna MS, Kendall PC. Computer-assisted CBT for child anxiety: the coping Cat CD-ROM. Cogn Behav Pract 2008 May;15(2):159-165. [doi: 10.1016/j.cbpra.2008.02.002]
60. Padgett DK. Qualitative Methods in Social Work Research: Challenges and Rewards. Newbury Park, CA: Sage Publications; 1998.
61. Lenhard F, Vigerland S, Engberg H, Hallberg A, Thermaenius H, Serlachius E. 'On my Own, but not Alone' - adolescents' experiences of internet-delivered cognitive behavior therapy for obsessive-compulsive disorder. PLoS One 2016;11(10):e0164311 [FREE Full text] [doi: 10.1371/journal.pone.0164311] [Medline: 27711249]
62. Scahill L, Riddle MA, McSwiggin-Hardin M, Ort SI, King RA, Goodman WK, et al. Children's Yale-Brown obsessive compulsive scale: reliability and validity. J Am Acad Child Adolesc Psychiatry 1997 Jun;36(6):844-852. [doi: 10.1097/00004583-199706000-00023] [Medline: 9183141]
63. Seidman I. Interviewing as Qualitative Research: A Guide for Researchers in Education and the Social Sciences. Third Edition. New York: Teachers College Press; 1991.
64. Schilling J. On the pragmatics of qualitative assessment. Eur J Psychol Assess 2006 Jan;22(1):28-37. [doi: 10.1027/1015-5759.22.1.28]
65. Kruger D, Swanepoel M. Gluing the pieces together: female adolescents' construction of meaning through digital metaphoric imagery in trauma therapy. Art Psychother 2017 Jul;54:92-104. [doi: 10.1016/j.aip.2017.04.011]

66. Grist R, Porter J, Stallard P. Acceptability, use, and safety of a mobile phone app (BlueIce) for young people who self-harm: qualitative study of service users' experience. JMIR Ment Health 2018 Feb 23;5(1):e16 [FREE Full text] [doi: 10.2196/mental.8779] [Medline: 29475823]
67. Nieto R, Hernández E, Boixadós M, Huguet A, Beneitez I, McGrath P. Testing the feasibility of DARWeb: an online intervention for children with functional abdominal pain and their parents. Clin J Pain 2015 Jun;31(6):493-503. [doi: 10.1097/AJP.0000000000000199] [Medline: 25551478]
68. Law EF, Beals-Erickson SE, Fisher E, Lang EA, Palermo TM. Components of effective cognitive-behavioral therapy for pediatric headache: a mixed methods approach. Clin Pract Pediatr Psychol 2017;5(4):376-391 [FREE Full text] [doi: 10.1037/cpp0000216] [Medline: 29503787]
69. Kaltenthaler E, Sutcliffe P, Parry G, Beverley C, Rees A, Ferriter M. The acceptability to patients of computerized cognitive behaviour therapy for depression: a systematic review. Psychol Med 2008 Nov;38(11):1521-1530. [doi: 10.1017/S0033291707002607] [Medline: 18205964]
70. Waller R, Gilbody S. Barriers to the uptake of computerized cognitive behavioural therapy: a systematic review of the quantitative and qualitative evidence. Psychol Med 2009 May;39(5):705-712. [doi: 10.1017/S0033291708004224] [Medline: 18812006]
71. McLeod BD, Wood JJ, Weisz JR. Examining the association between parenting and childhood anxiety: a meta-analysis. Clin Psychol Rev 2007 Mar;27(2):155-172. [doi: 10.1016/j.cpr.2006.09.002] [Medline: 17112647]
72. de Graaf LE, Gerhards SA, Arntz A, Riper H, Metsemakers JF, Evers SM, et al. One-year follow-up results of unsupported online computerized cognitive behavioural therapy for depression in primary care: a randomized trial. J Behav Ther Exp Psychiatry 2011 Mar;42(1):89-95. [doi: 10.1016/j.jbtep.2010.07.003] [Medline: 20723885]
73. Bandura A. Social learning of moral judgments. J Pers Soc Psychol 1969 Mar;11(3):275-279. [doi: 10.1037/h0026998] [Medline: 5784267]

Abbreviations

CBT: cognitive behavioral therapy
cCBT: computerized cognitive behavioral therapy
HCI: human-computer interaction
RCT: randomized controlled trial
SPARX: smart, positive, active, realistic, X-factor thoughts
TEAM: Technology Enabled Mental Health for Young People
tech-assisted CBT: technology-assisted cognitive behavioral therapy
TFBs: thoughts, feelings, and behaviors

Edited by G Eysenbach; submitted 30.01.19; peer-reviewed by K Stasiak, M Birk; comments to author 18.04.19; revised version received 12.05.19; accepted 29.05.19; published 12.11.19.

Please cite as:
McCashin D, Coyle D, O'Reilly G
Qualitative Synthesis of Young People's Experiences With Technology-Assisted Cognitive Behavioral Therapy: Systematic Review
J Med Internet Res 2019;21(11):e13540
URL: http://www.jmir.org/2019/11/e13540/
doi:10.2196/13540
PMID:31714251

Original Paper

Evaluating the Efficacy of Internet-Delivered Cognitive Behavioral Therapy Blended With Synchronous Chat Sessions to Treat Adolescent Depression: Randomized Controlled Trial

Naira Topooco[1,2], PhD; Sandra Byléhn[1], MSc; Ellen Dahlström Nysäter[1], MSc; Jenny Holmlund[1], MSc; Johanna Lindegaard[1], MSc; Sanna Johansson[1], MSc; Linnea Åberg[1], MSc; Lise Bergman Nordgren[3], PhD; Maria Zetterqvist[4,5], PhD; Gerhard Andersson[1,3], PhD

[1]Department of Behavioural Sciences and Learning, Linköping University, Linköping, Sweden

[2]Center for m[2]Health, Palo Alto, CA, United States

[3]Centre for Psychiatry Research, Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden

[4]Center for Social and Affective Neuroscience, Department of Clinical and Experimental Medicine, Linköping University, Linköping, Sweden

[5]Department of Child and Adolescent Psychiatry, Region Östergötland, Linköping, Sweden

Corresponding Author:
Naira Topooco, PhD
Department of Behavioural Sciences and Learning
Linköping University
Campus Valla
Linköping, SE-581 83
Sweden
Phone: 46 13281000
Email: naira.topooco@liu.se

Abstract

Background: Depression is a common and serious problem among adolescents, but few seek or have access to therapy. Internet-delivered cognitive behavioral therapies (ICBTs), developed to increase treatment access, show promise in reducing depression. The inclusion of coach support in treatment is desired and may be needed.

Objective: The aim of this study was to determine the efficacy of an ICBT protocol blended with weekly real-time therapist sessions via chat; blended treatment, for adolescent depression, including major depressive episode (MDE). The protocol has previously been evaluated in a controlled study.

Methods: In a two-arm randomized controlled trial, adolescents 15 to 19 years of age were recruited through a community setting at the national level in Sweden (n=70) and allocated to either 8 weeks of treatment or to minimal attention control. Depression was assessed at baseline, at posttreatment, and at 12 months following treatment (in the intervention group). The primary outcome was self-reported depression level as measured with the Beck Depression Inventory II at posttreatment. The intervention was offered without the need for parental consent.

Results: Over two weeks, 162 adolescents registered and completed the baseline screening. Eligible participants (n=70) were on average 17.5 years of age (SD 1.15), female (96%, 67/70), suffered from MDE (76%, 53/70), had no previous treatment experience (64%, 45/70), and reported guardian(s) to be aware about their depression state (71%, 50/70). The average intervention completion was 74% (11.8 of 16 modules and sessions). Following the treatment, ICBT participants demonstrated a significant decrease in depression symptoms compared with controls (P<.001), corresponding to a large between-group effect (intention-to-treat analysis: d=0.86, 95% CI 0.37-1.35; of completer analysis: d=0.99, 95% CI 0.48-1.51). A significant between-group effect was observed in the secondary depression outcome (P=.003); clinically significant improvement was found in 46% (16/35) of ICBT participants compared with 11% (4/35) in the control group (P=.001).

Conclusions: The results are in line with our previous study, further demonstrating that adolescents with depression can successfully be engaged in and experience significant improvement following ICBT blended with therapist chat sessions. Findings on participants' age and baseline depression severity are of interest in relation to used study methods.

Trial Registration: ClinicalTrials.gov NCT02363205; https://clinicaltrials.gov/ct2/show/NCT02363205

(J Med Internet Res 2019;21(11):e13393) doi:10.2196/13393

KEYWORDS

adolescent; depression; cognitive behavioral therapy; randomized controlled trial; internet; digital health; technology; mental health; text messaging; instant messaging

Introduction

Background

Unipolar depressive disorders are the leading cause of disability-adjusted life years among adolescents aged 15 to 19 years globally [1]. Early age of onset of depression is a risk factor for recurrent depression [2,3], is associated with poor academic achievement, unemployment, and impaired quality of life [2,4-6] and predicts additional and worsened mental and physical illness [7-9]. Despite the high level of disability, only a small fraction of young people in need have access to any kind of intervention [10,11].

Internet-Delivered Psychological Treatment

At this time, psychological treatment in the form of cognitive behavioral therapy (CBT) is considered one of the best empirically supported behavioral interventions to reduce depression [12-15]. The theoretical framework of CBT is rooted in the central assumption that depression is caused and maintained by unhelpful cognitions and behaviors, with treatment accordingly focused on improving function in these domains with the application of skill-based behavioral strategies [16,17]. Worldwide, adaptions of in-person psychotherapy protocols into Web-based formats, the majority involving CBT, are emerging to bring mental health interventions to individuals who for different reasons are not reached by regular services. In internet-delivered CBT, often referred to as ICBT, content in the form of text, video, or audio is arranged into weekly *modules* or *sessions* and delivered inside a Web-based treatment platform along with homework assignments [18]. Programs are offered as self-help or involve a clinician with the role of guiding the participant through the program and providing brief feedback on treatment progress. Clinician support is often provided asynchronously (eg, email), with the administration time typically not extending 20 min per participant and week [19].

The Role of Support in Internet-Delivered Cognitive Behavioral Therapy

Concerning ICBT and other computer and Web-based interventions for depression, one of the most consistent findings is that human support matters. The social element—having someone to talk to or just knowing that someone is there monitoring and listening, seems important in maintaining motivation to continue treatment and experiencing improvement from it. Although there are exceptions in the literature (eg [20]), coached programs are shown to be more effective than self-guided programs in the treatment of adult depression [21-23] with guided ICBT demonstrating effects similar to those found with in-person CBT [24,25]. In comparison, self-guided ICBTs produce small treatment effects that sometimes merely surpass the lower cut-off point for what is considered clinical relevance in the treatment of depression [26,27]. *True* self-help ICBTs, delivered freely to anyone and without any human interaction in connection to the intervention (eg, interview and administrative contact), show among the highest dropout rates and smallest effects in the field [28]. A study on patients' perspectives on ICBT for depression points out the importance of guidance in making treatment work [29]. Existing ICBT programs produce benefits for youth depression, but there is considerable variation in populations, measures used, and outcomes [11,30-32]. The influence of support on outcomes in computer and Web-based intervention for youth depression and anxiety, however, was recently reported in a meta-analysis by Grist et al [33], in line with results for ICBT with adult populations. On the basis of 34 studies with 3113 children and adolescents, programs that included >90 min supportive contact yielded higher effect sizes (g=0.87) than purely self-help programs (g=0.24), although interventions based on theoretical frameworks other than CBT were included in the analyses. A review focusing on how Web-based depression programs for adolescents work, including ICBT programs, found that completion rates increased if the treatment was delivered with real-time guidance from a doctor, therapist, or teacher [34]. Even programs that included automated reminders, praise, or suggestions were only able to optimize adherence with real-time, in-person contact. On the individual study level, depression prevention studies with youth have found an almost 10-fold difference in program completion between ICBT when offered with teacher support or monitoring as opposed to self-help [35], and that adolescents report stopping intervention because of the need to talk to someone, rather than doing a program [36]. It is not necessarily a problem that unguided behavioral interventions offered to communities are associated with limited effects—no additional cost is associated with repeated use of the intervention, and despite, for example, high dropout rate, a large number of individuals will still potentially benefit from the interventions. However, for the treatment of clinical depression and in the treatment of youth—a particularly vulnerable population—the apparent ability of coach support to improve outcomes in ICBT warrants consideration of how support can be strengthened and provided to adolescents.

A Text-Based Blended Treatment Approach

To include a strong therapist interaction in the form of sessions is in line with aggregated findings that support is desired and boosts the effect of ICBT. Indeed, emerging *blended treatment* approaches [37,38] that integrate some in-person therapist sessions with mobile or internet sessions in the same protocol—with the rationale of providing a strong therapist interaction while keeping the advantage of reliable self-help components—show promise in outcomes [39,40] and acceptability [41,42] for adult depression. Among young people, the preference for real-time texting and instant messaging (chat) is well established [43,44]. In the United States, nearly 1 in 3 (32%) of adolescents and young adults with moderate to severe depressive symptoms report having used texting, Web-based

messaging, an app, or video chat to connect with a health care provider [45]. To include strong therapist support in the form of sessions (ie, blended treatment), and moreover, doing so with real-time texting is consistent with adolescents' media use preferences, while maintaining the combination of easy access and discretion offered in ICBT, which for young people may prove especially effective in overcoming barriers to behavioral intervention. The inclusion of strong therapist interaction could moreover extend ICBT to appropriately address clinical depression in youth. In light of full-threshold depressive episodes being commonly experienced in adolescence and young adulthood despite prevention effort [46], this is highly relevant.

Objective

This study investigated a treatment protocol consisting of ICBT modules blended with weekly therapist chat sessions for adolescent depression, including major depressive episode (MDE). We have previously evaluated the treatment in a controlled trial with promising results (d=0.71 against minimal attention control [47]). In line with participant feedback, the protocol was subsequently revised to include longer sessions while the conceptual model of delivery was kept intact. The main objective of this study was to further establish the effect of the treatment model, using comparable eligibility criteria and study methods, including allowing adolescents aged 15 to 17 years to participate without parental involvement. We hypothesized that the intervention would outperform the control condition in reducing depression, corresponding to at least moderate between-group effect.

Methods

Study Design and Participants

In a 2-arm randomized controlled trial, adolescents were recruited in a community setting at the national level in Sweden and randomized (1:1 ratio) to ICBT (n=35) or to minimal attention control (n=35). Enrollment and baseline assessments took place between January 26 and February 10, treatments were conducted between February 13 and April 9, and posttreatment assessments (at 8 weeks) were conducted between April 10 and 16, 2017. Controls were given access to treatment following the posttreatment assessment (8 weeks), and ICBT participants were reassessed 12 months following treatment (April 2018). Eligible participants were 15 to 19 years, suffered depressive symptoms (≥14 points on the Beck Depression Inventory II; BDI-II [48]), and presented at least 4 symptoms including 1 core symptom, or fulfilled criteria for major depressive episode (MDE) according to The Mini-International Neuropsychiatric Interview (MINI 7.0 [49]). We excluded individuals who were receiving psychological therapy, were alcohol or drug dependent, showed severe suicidal ideation, or who had severe comorbid psychiatric conditions (eg, bipolar disorder or psychotic symptoms). Comorbid anxiety disorder(s) were allowed if depression was the principle concern. Medication for anxiety, depression, or Attention deficit hyperactivity disorder was accepted if the dose had been stable >1 month before the study. Minors aged 15 to 17 years were included in line with Swedish research legislation, which states that individuals aged 15 to 17 years can consent to research without parental involvement if possessing sufficient maturity to freely undertake study participation with awareness of possible adverse consequences.

Participants were recruited by social media posts from study staff including one guest post in a wide-reaching Instagram account focusing on coping with mental health issues. Posts described the opportunity to receive Web-based psychological treatment within a research study. Information was also distributed to schools, youth centers, and clinics across Sweden. Potential participants were directed to the study website and registered for the study by creating a user account, providing informed consent (checking a box), and completing a Web-based screen (demographics and self-reported outcome measures). An encrypted Web-based treatment platform, *Iterapi*, was used to collect screening data [50]. Individuals who showed initial eligibility were invited to a phone interview with study staff to confirm eligibility using the full MINI, to determine matureness to participate, to obtain verbal consent, and to confirm identity (name, address, and personal identity number). The principal investigator decided on final eligibility. Before the randomization, eligible participants were requested to sign a digital consent sheet (full name with digital date stamp) to confirm their willingness to participate in the study. After consent was agreed and baseline data collected, participants were stratified according to depression severity (fulfilling DSM-5 criteria for MDE or not), and thereafter randomized in a 1:1 ratio. A person not involved in the study executed the randomization procedure using a computer-generated sequence service. Treatment was given open label. The Research Ethics Board in Linköping, Sweden, gave approval for the study (Reg. no. 2014/427-31). The study was registered at ClinicalTrials.gov (NCT02363205). Participants were not offered financial compensation for treatment or assessment completion at any time.

Interventions: Internet-Delivered Cognitive Behavioral Therapy Program

The treatment included 8 ICBT modules, and 8 individual therapist sessions delivered via chat; the entire intervention lasted 8 weeks. Treatment took place within an encrypted Web-based treatment platform: *Iterapi* [50]. Modules comprised text material and videos, fictional storylines, reflection tasks and homework assignments, and entailed the behavioral and cognitive approach of CBT. Core techniques included *behavioral activation:* detecting unhelpful behavior, and reinstating and reinforcing behaviors that increase positive consequences, thus elevating mood [51,52], and *cognitive restructuring*: correcting maladaptive thinking patterns and inaccurate beliefs (negatively biased views of oneself, of the world in general, and of the future) to reduce depression [53]. Table 1 presents an overview of modules. Therapist chat sessions were coscheduled by participant and therapist each week and were conducted inside the treatment platform. Sessions dealt with the previous and current module and focused on process-related aspects of treatment: identifying problems, examining the patient's cognitions, encouragement, answering questions, and assisting with homework assignments. Participants who agreed were sent reminders before sessions, those who missed a session were offered a new time on the basis

of their therapists' remaining availability for the week. In addition to sessions, therapists responded to homework assignments and questions within 24 h on weekdays. The treatment was to be completed on a preset pace of 8 weeks, thereafter participants had access to the program for 4 additional weeks without therapist support.

Compared with our previous study [47], the treatment protocol was revised according to participants feedback: sessions were prolonged from 30 to 45 min, texts in some modules were revised to present more clearly (eg, reworked sentences, a more clear description of the rationale for exposure), and information procedures given to participants about weekly themes and goals were standardized. The overall conceptual model of delivery was kept intact.

Controls were assigned to a therapist and received an introductory personal platform in-mail from their therapist, informing them that there would be weekly assessments and that their assessments were to be viewed by their therapist to monitor their mental health state. They were informed that their therapist might contact them to follow-up on their wellbeing. While being in the control group, participants were allowed to seek regular care, which in Sweden is for free for adolescents.

Table 1. Internet-delivered cognitive behavior therapy intervention overview.

Week	Web-based session	Assignment/exercise
1	Psychoeducation	Write history, set goals
2	Analysis of behavior	Identify dysfunctional and functional schemas
3	Behavioral activation	Mood-activity diary
4	Behavioral activation	Mood-activity diary
5	Cognitive restructuring	Identify and challenge thoughts
6	Psychoeducation (anxiety)	Anxiety management, graded exposure
7	Emotional recognition	Coping strategies, self-esteem, affect regulation
8	Maintenance	Relapse prevention, treatment summary

Therapists and Safety Procedures

In total, 6 CBT *therapists* in training conducted study assessments and treated 5 to 6 ICBT participants and monitored 5 to 6 control participants each. Before assessments, therapists received training in clinician interview (½ day), ICBT (written material), and how to navigate the treatment platform (½ day, on demand). Therapists received 60 min of clinical supervision on a weekly basis from clinical psychologists with expertise in adolescent psychopathology and delivery of ICBT. Communication and records (eg, chat, platform messages, and study consent) were available for the participant and therapist to view at any time. A short version of the Mood and Feelings Questionnaire (MFQ-13) and the suicidal ideation item from the Patient Health Questionnaire were used for weekly monitoring of depression. Participants were instructed to immediately contact their therapist in the event of feeling worse and were informed that their therapist might contact them in case of noncompletion. Scores and messages were monitored on a daily basis. In cases of suicidal ideation or significant deterioration, participants were immediately followed up by email and phone. The study collected participants personal identity number and address, and informed participants that in the event of imminent crisis, the study would break confidentiality to pursue appropriate follow-up.

Outcomes

The primary outcome was self-reported depression severity at posttreatment, measured by the BDI-II [48]. Secondary outcomes included the MFQ [54]; the beck anxiety inventory (BAI) [55]; the social interaction anxiety scale (SIAS) [56]; the general self-efficacy scale (GSE) [57]; the credibility expectancy questionnaire [58]; the working alliance inventory (WAI-S) [59] and the Brunnsviken Brief Quality of life scale (BBQ) [60]; all self-report scales were completed over Web. The MINI was readministered over phone at posttreatment to assess depression diagnosis, assessors were not blinded to participant allocation at posttreatment.

Analyses

Our previous study [47] was used as a reference for power calculations. To detect a similar effect size (Cohen d=0.70) at posttreatment, with a 2-tailed 5% significance level and a power of 80%, a total sample size of 72 was required. *A priori:* Participants were included in statistical analyses according to the intention-to-treat (ITT) principle. Missing data were handled using multiple imputations. Differences in primary outcomes were evaluated pre- to posttreatment by analysis of covariance (ANCOVA) using baseline values as covariate at the P<.05 level. Effect sizes were calculated based on imputed values and observed standard deviations. *Post hoc*: Independent *t* tests and Pearson chi-square tests were used to detect possible baseline differences between groups and percentage decrease in symptoms, respectively. Little missing completely at random test was performed to test the assumption of data missing at random, Levene test was performed to test the assumption of equal variance between groups. Completers were included in complementary statistical analyses for the primary and secondary outcomes. Clinically significant change [61] was determined by investigating the number of participants falling *2 SD* below the pretreatment mean for both conditions on the primary outcome, while fulfilling The reliable change index criteria [62], a psychometric criterion used to determine whether an individual change score between baseline and posttreatment assessment is significantly greater than a difference that could have occurred because of random measurement error. A criterion

of 30% or more increase on the primary outcome from baseline to posttreatment was used to determine significant deterioration.

Results

Participant Flow

Figure 1 presents a Consolidated Standards of Reporting Trials diagram of the participant flow through the study. During 2 weeks of registration, 162 individuals completed the initial Web-based screening, out of whom, 12 identified as males. For included participants, the mean time from initial screening to the initiation of treatment or control was 12.7 days (range 4-17 days). No significant between-group differences were found at baseline regarding demographics or outcome measures. Table 2 presents the study sample baseline characteristics. During the treatment period, the average completion of the weekly assessment (both allocations) was 93% (range 84%-100%). Table 3 presents completion rates for primary and secondary outcomes at posttreatment. Post hoc analyses showed no differences between those lost to posttreatment and the rest of the sample on any outcome measures at baseline.

Figure 1. Participants' flow through the study. MINI: Mini-International Neuropsychiatric Interview; ICBT: internet-delivered cognitive behavioral therapy; BDI-II: Beck Depression Inventory-II.

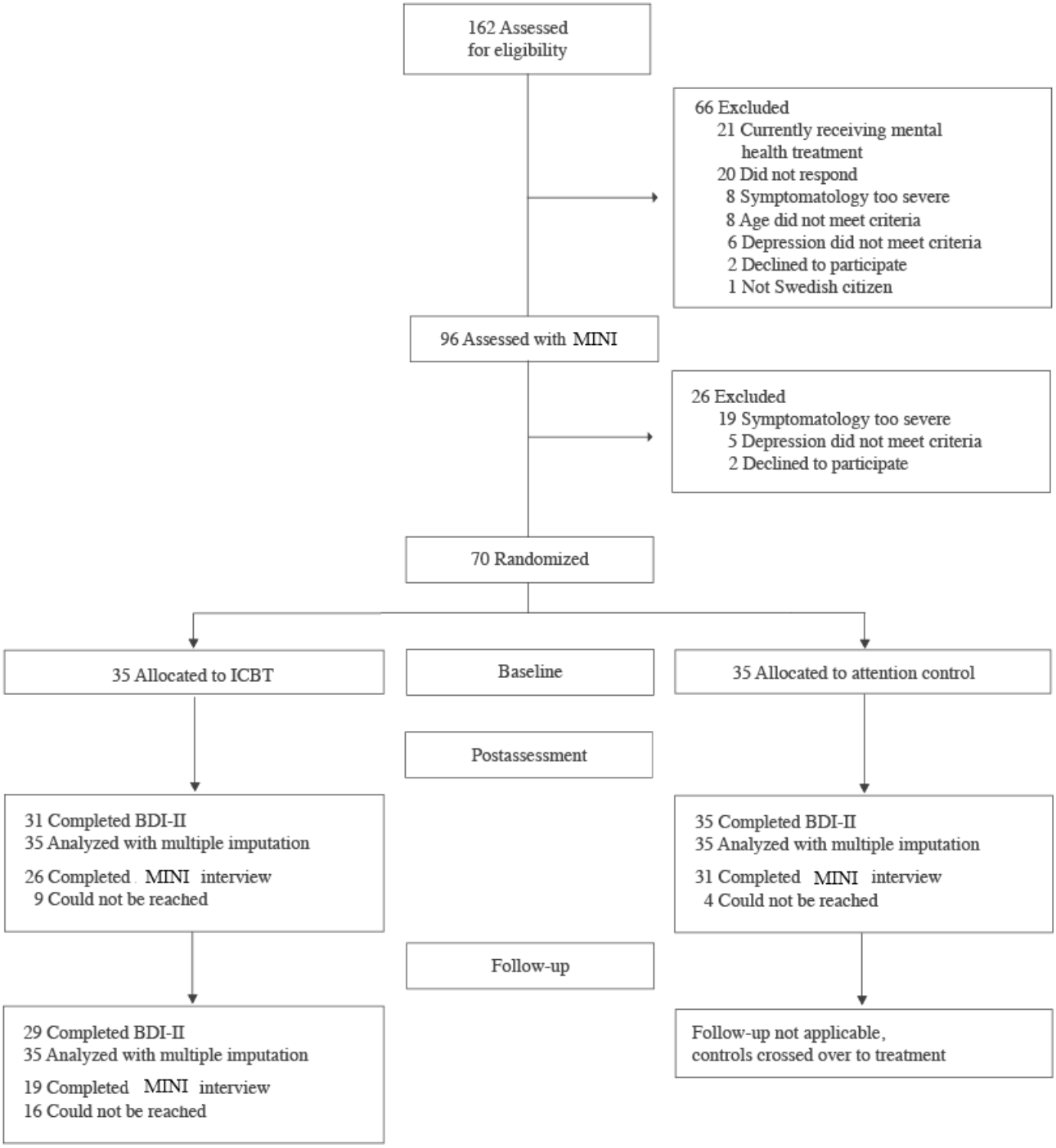


Table 2. Baseline characteristics of participants.

Characteristics	ICBT[a] (n=35)	Control (n=35)
Female, n (%)	32 (91)	35 (100)
Age, mean (SD)	17.5 (1.1)	17.5 (1.2)
Occupation, n (%)		
Studying full time	28 (80)	32 (91)
Hiatus/dropout	2 (6)	0 (0)
Working	5 (14)	3 (9)
Residence, n (%)		
City	10 (29)	7 (20)
Small town/country side	25 (71)	28 (80)
Family, n (%)		
Two-parent household	17 (49)	9 (26)
Other	18 (51)	26 (74)
Parent(s) country of birth other than Sweden	7 (20)	7 (20)
Major depressive episode[b], n (%)	27 (77)	26 (74)
18-19 years	14 (74)	16 (80)
15-17 years	13 (81)	10 (67)
Comorbid anxiety diagnosis, n (%)		
Any	25 (71)	24 (69)
Generalized anxiety disorder	14 (40)	15 (43)
Social anxiety disorder	16 (46)	10 (29)
Panic disorder	11 (31)	11 (31)
Agoraphobia	4 (11)	7 (20)
Guardians informed about mental health state, n (%)	25 (71)	25 (71)
Previous treatment history, n (%)	10[c] (29)	15[d] (43)
Psychotherapy treatment	10 (29)	14 (40)
Psychotropic medication	1 (3)	4 (11)
Current treatment, n (%)	7[e] (20)	3 (9)
Counselor support	3 (9)	1 (3)
Psychotropic medication	5 (14)	2 (6)

[a]ICBT: internet-delivered cognitive behavioral therapy.

[b]Confirmed in The Mini-International Neuropsychiatric Interview 7.0.

[c]One participant had experience of psychotherapy treatment as well as psychotropic treatment, thus total n=10.

[d]Some participants had experience of both types of treatment.

[e]Some participants had current experience of support as well as psychotropic medication.

Table 3. Participants' assessment completion.

Measure completed	Posttreatment		12 months	
	ICBT[a]	Control	ICBT	Control
BDI-II[b], n (%)	31 (89)	35 (100)	29 (83)	—[c]
BBQ[d], n (%)	31 (89)	35 (100)	28 (80)	—
MFQ[e], n (%)	31 (89)	34 (97)	28 (80)	—
BAI[f], n (%)	31 (89)	34 (97)	28 (80)	—
SIAS[g], n (%)	31 (89)	34 (97)	28 (80)	—
GSE[h], n (%)	31 (89)	34 (97)	27 (77)	—

[a]ICBT: internet-delivered cognitive behavioral therapy.

[b]BDI-II: Beck Depression Inventory II.

[c]Data not applicable.

[d]BBQ: Brunnsviken Brief Quality of Life Inventory.

[e]MFQ: Mood and Feelings Questionnaire.

[f]BAI: Beck Anxiety Inventory.

[g]SIAS: Social Interaction Anxiety Scale.

[h]GSE: General Self-Efficacy scale.

Primary Outcome

Table 4 presents pre and posttreatment assessments including effect sizes, means, and standard deviations for both groups, at pre and posttreatment. For the primary outcome measure BDI-II, analyses with ANCOVA with baseline scores as covariate revealed a significant effect between groups ($F_{1,67}$=22.23, P<.001) at posttreatment. The corresponding between-group effect size was d=0.86 (95% CI 0.37-1.35). A post hoc completer analysis (n=66) showed a similar result (d=0.99, 95% CI 0.48-1.51; P<.001). Figure 2 illustrates within-group improvements on the BDI-II. The ICBT group was reassessed 12 months following treatment. Analysis with paired t test from posttreatment to follow-up (ITT analysis) indicated that there was no difference in depression level from posttreatment assessment to the follow-up (BDI-II, P=.96).

Table 4. Means (SD) and effect sizes (Cohen *d*) with 95% CI for continuous outcome variables, with missing data imputed.

Measure	Baseline, mean (SD)	Posttest, mean (SD)	12 Months, mean (SD)	Cohen *d* between-group posttest (95% CI)
BDI-II[a]				
ICBT[b]	31.6 (10.0)	16.0 (11.3)	15.9 (16.1)	0.86 (0.37 to 1.35)[c]
Control	28.8 (7.9)	24.8 (10.4)	—[d]	—
MFQ[e]				
ICBT	36.0 (10.7)	24.3 (12.8)	21.7 (17.4)	0.58 (0.10 to 1.06)[f]
Control	35.2 (9.4)	31.0 (9.8)	—	—
BBQ[g]				
ICBT	35.8 (18.1)	46.7 (21.3)	48.3 (27.0)	0.34 (0.19 to 1.15)[f]
Control	38.7 (17.2)	39.1 (15.7)	—	—
BAI[h]				
ICBT	28.6 (11.9)	16.6 (10.3)	15.8 (12.7)	0.30 (–0.17 to 0.77)
Control	25.5 (11.2)	20.0 (9.3)	—	—
SIAS[i]				
ICBT	45.2 (19.2)	35.4 (19.0)	37.4 (22.9)	0.05 (–0.41 to 0.52)
Control	39.5 (16.4)	35.1 (14.3)	—	—
GSE[j]				
ICBT	21.3 (5.6)	22.9 (7.5)	24.3 (9.6)	0.10 (–0.37 to 0.56)
Control	22.2 (4.6)	23.0 (5.0)	—	—

[a]BDI-II: Beck Depression Inventory II.

[b]ICBT: internet-delivered cognitive behavioral therapy.

[c]$P<.001$.

[d]Data not applicable.

[e]MFQ: Mood and Feelings Questionnaire.

[f]$P<.01$.

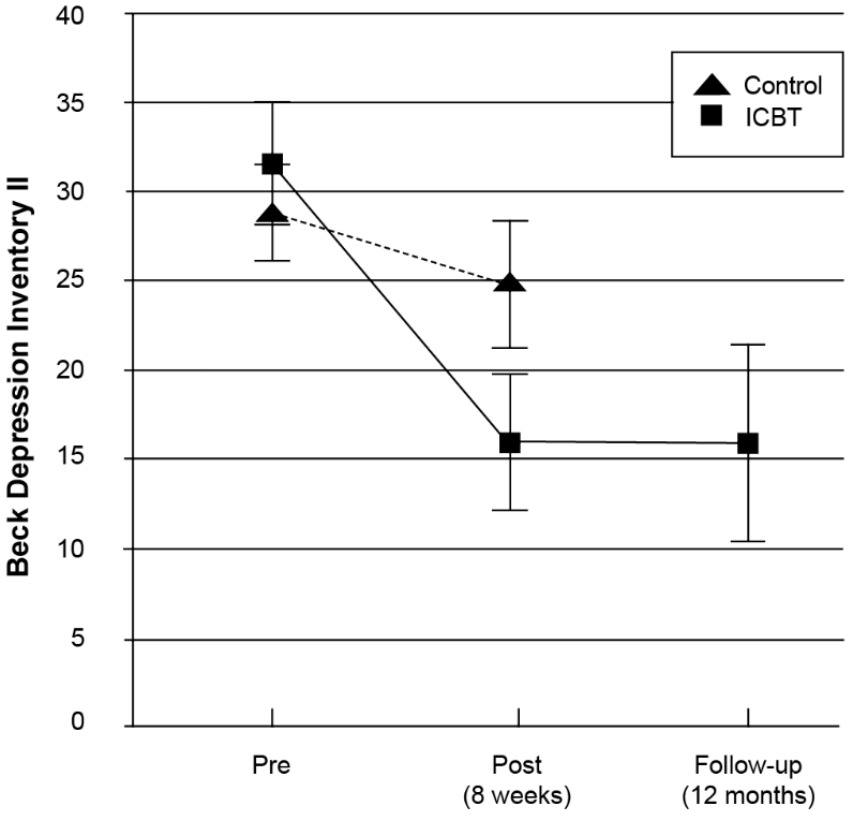
[g]BBQ: Brunnsviken Brief Quality of Life Inventory.

[h]BAI: Beck Anxiety Inventory.

[i]SIAS: Social Interaction Anxiety Scale.

[j]GSE: General Self-Efficacy Scale.

Figure 2. Change over time in depression severity (95% CIs) at baseline, posttreatment, and 12-month follow-up. ICBT: internet-delivered cognitive behavioral therapy.



Response and Remission

We performed post hoc analyses to further investigate potential change in depression symptom level, with missing cases (ICBT n=4) categorized as not improved. At posttreatment, a higher proportion of ICBT participants (46%, 16/35) than controls (11%, 4/35) showed clinically significant improvement ($\chi^2_{1,70}$=10.1; *P*=.001), defined as scoring *2 SD* below the pretreatment mean for both conditions on the BDI-II, while also fulfilling the reliable change index criteria [62]. In addition, the number of participants presenting a decrease of ≥30% in BDI-II score from baseline to posttreatment assessment, and BDI-II ≥13, and BDI-II ≥10 at postassessment was investigated as these measures have previously been used to define significant changes or cut-offs [48]. Significant differences between groups were found for all measures (*P*=.004, *P*=.004, *P*<.001). Participants were reassessed with the MINI at posttreatment. Of those participants that had fulfilled DSM-5 criteria for MDE at baseline (ICBT n=27; Control n=26), a higher proportion of ICBT participants (56%, 15/27) than controls (27%, 7/26) no longer met diagnostic criteria at posttreatment ($\chi^2_{1,53}$=2.0; *P*=.03); missing cases were categorized as not improved (ICBT n=7; Control n=4).

Secondary Outcomes

At posttreatment, a significant effect was found for the secondary depression outcome MFQ ($F_{1,67}$=9.33; *P*=.003), corresponding to a medium between-group effect of *d*=0.58 (95% CI 0.10-1.06). A post hoc completer analysis (n=66) showed a similar result (*d*=0.67; 95% CI 0.18-1.17; *P*=.002). Analysis with paired *t* test from posttreatment to follow-up (ITT analysis) indicated that there was no difference in depression level from posttreatment assessment to the follow-up (MFQ, *P*=.25). The ANCOVA for posttreatment change in quality of life (BBQ) revealed a significant effect between groups ($F_{1,67}$=8.73; *P*=.004). The ANCOVA for pre to posttreatment change in anxiety (BAI) showed no significant effect between groups ($F_{1,67}$=3.95; *P*=.051), nor did the ANCOVAs for self-efficacy (GSE; *P*=.81) or social anxiety (SIAS, *P*=.86).

Program Use

ICBT participants logged into the treatment platform for a mean of 28.4 times (SD 14.6) and completed on average 78% of available modules (mean 6.2 of 8 modules, SD 2.28), and on average 71% of available therapist sessions (mean 5.7 of 8 sessions, SD 2.67). The average total completion was 74% of all 16 available sessions and modules (11.8/16, SD 4.82). In total, 17% participants (6/35) completed less than half of the available treatment modules. No relationship was found between the number of completed sessions and modules and treatment outcome (*P*=.10). Therapists spent on average 43.6 min (SD 19.4) on each participant every week (if only including recordings that included therapist sessions, mean 55.8 min on each participant every week). ICBT participants' ratings of treatment credibility (C-scale) showed an average rating of 18.5 (SD 4.17) out of a maximum total of 27 (highest credibility). The average score on items of therapeutic alliance (WAI-S) was 4.95 (SD 0.63) out of a maximum of 7 (highest satisfaction).

Negative Outcomes

One participant in the ICBT group deteriorated significantly during the course of treatment and was directed to standard care services while being maintained in the study. Post hoc analyses showed that no participant deteriorated significantly following treatment, defined as an increase of 30% or more on the BDI-II from baseline to posttreatment. If those lost to posttreatment assessment were categorized as having deteriorated (n=4), the rate in the ICBT group would be 11%.

Discussion

We examined whether ICBT blended with weekly synchronous therapist chat sessions was effective in reducing depression when compared with minimal attention control. The intervention was evaluated with adolescents 15 to 19 years of age (mean 17.5 years) suffering from depressive symptoms, including but not restricted to, MDE. Our report demonstrates the efficacy of the intervention and highlights the potential of a Web, text-based CBT approach to reach and treat adolescents suffering depression.

Principal Findings

We found superiority for the intervention based on the primary outcome measure (BDI-II score at week 8), corresponding to a large between-group effect size (d=0.86). Significant effects on depression symptom level were also observed in the secondary self-reported depression outcome (MFQ, d=0.58), as well as in the number of cases in remission from a major depressive episode (determined using the MINI), further supporting that the intervention was effective. At the 12-month follow-up, depression levels in the treatment group were similar to those observed at posttreatment, follow-up data were not available for the control group. The results for the depression outcomes are in line with, and for the primary outcome surpass, the findings in our previous study that examined the intervention [47]. Although it is important to consider that we made changes to the protocol between studies—foremost, session time was prolonged—the overall conceptual model of delivery was kept intact, and we used comparable design, measures, and procedures across studies. We therefore consider that the findings in this study provide further support for the efficacy of ICBT modules blended with chat sessions to reduce adolescent depression.

Relation to Previous Research

Our protocol included more therapist support than typically found in ICBT. We learned in our previous study that therapist sessions were utilized (completion, mean 78%) but perceived short, and in line with participants' feedback, we prolonged session time in this study. Sessions were completed at a similar degree in this study (mean 71%) with fewer comments on sessions length. Possibly, the extension in time resulted in chat sessions being perceived as sufficient. Modules and sessions were completed at a similar degree, and few participants completed less than half of the treatment. This may explain why we found no association between dose and depression outcomes. Our tentative conclusion, based on participant feedback, program completion, and findings with the therapist work-alliance (WAI-S), is that real-time therapist interaction is indeed desired when offered in tandem with ICBT material, and that completely text-based sessions seem to generate a meaningful therapeutic relationship for young individuals. We treated participants with high depression severity, and it may be that an ICBT format with strong therapist support, *blended treatment*, can better address clinical levels of depression than highly automated interventions do, because blended treatment can be tailored to suit individual needs, help with motivation, and provide more opportunity to monitor signs of improvement and deterioration, respectively [37,63]. In this context, Sethi et al have previously reported a blended approach, consisting of Web-based ICBT and in-person sessions, to be superior to stand-alone Web-based ICBT, and similar to in-person CBT in reducing depression in young adults (mean 19.5/20.1 years [64,65]). Chat sessions in group format to deliver ICBT with adolescents and young adults (mean 20.9 years) has previously been evaluated in a study by van der Zanden et al [66]. They too reported a large effect size on depressive symptoms against waitlist at posttreatment (d=0.94), and similar rate of clinically reliable change (56%) as in this study.

On a general level, the completion rate in this study compares favorably with what is reported for Web-based interventions [67], including for youth [33], and it should be considered that we did not offer any payment for completion or assessment. The depression reduction observed in this study is moreover in line with meta-analyses suggesting that computer and Web-based ICBT for youth depression produces effect [11,32], and more broadly, that coached Web-based interventions fare better than self-guided [33]. Existing studies, however, vary considerable in relation to the populations, delivery mode, and measures in focus, making comparison difficult. For example, our obtained effect size is in line with the larger effect size found for computer-based and Web-based CBT with adolescents aged >13 years against waitlist (g=0.95), as compared with children (g=0.51), reported in a meta-analysis subanalysis [11], but the authors did not report interventions for anxiety and depression programs separately or differentiated on level of coach support. Most of previous RCT's that have focused on elevated depression in adolescents and young adults [68-78] have evaluated ICBT self-guided programs [68-72]; telephone calls could occur [69], or interventions that are not ICBT models, but rather computerized programs (eg, fantasy games) with support or oversight at education site, school, or at home [73-78]. To the best of our knowledge, no previous ICBT programs have involved individual therapist sessions with adolescents with current major depressive episode. Although our results are promising and could be associated with the level of therapist support, this was not established by the methods used, and we cannot rule out that other factors have played a role, for example, time to treatment, and initial depression severity. This should be addressed in future studies.

Clinical Implications

This study focused on efficacy, nevertheless findings may in several aspects have clinical implication. First, we used a recruitment strategy that is similar to the way in which some young individuals enroll for Web-based mental health intervention [79], thus giving ecological validity to this study

and the population with which it is conducted. Second, our findings support that reducing adolescent depression, including major depressive episode, is possible using a genuinely Web-based CBT approach. Third, our results compare favorably with those reported in meta-analyses on in-person CBT with youth depression as compared with active and nonactive control conditions [80-82], and more broadly, with effects for in-person psychotherapies for youth depression (g=0.29) [83]. Fourth, observed diagnosis remission rates are in the lower range compared with studies that have reported remission for in-person CBT: 61% to 87% [84-86]. Participants in our study had high depression scores initially, with large reductions needed to reach the criteria for remission. It has been suggested that young people who access mental health resources over Web are likely to be in greater distress in comparison with those accessing in-person services (given more rapid access to Web, they are closer in time to their symptoms [87]), which point toward the relevance of Web-based behavioral interventions extending to address clinical depression. Fifth, there are concerns among stakeholders about the safety and effectiveness of internet interventions for depression, particularly regarding offering such interventions to youth [88,89], which is why this study reports on potential negative effects related to intervention. We found 1 participant deteriorating significantly during the course of treatment. In comparison, it has been reported that 14% to 24% of young people receiving psychotherapy in US outpatient mental health services experience significant deterioration following treatment [90]. Concerning ICBT, a review reports 5.8% of adult study participants experiencing significant deterioration following self-guided programs [91]; no review has investigated youth. Sixth, in contrast to clinical practice [92] and research in many places, this study did not require participants to visit a care facility or school counselor or inform their guardians to receive treatment. This was to acknowledge that some adolescents do not want to, or cannot involve, their parents in their difficulties [93], and that this reluctance may delay or impede evaluation and treatment [94]. We found the intervention to attract young individuals (mean 17.5 years), some reporting that they had not informed parent(s) about their mental state (29%). The findings on participants' characteristics, as well as the other findings reported on here, are very similar to those in our previous study [47], and further contribute to reveal who may be reached and treated when similar interventions are offered. If the aim is to treat young individuals, at an early stage of the disorder, the conditions required to receive treatment may benefit from becoming more inclusive.

Strengths and Limitations

Strengths of the study include a rigorous design, use of a primary outcome with strong psychometric evidence, and adequate power with regards to detecting between-group differences in the included depression outcomes. In relation to the reproducibility challenges in psychology, it is positive that our results are in line with those in our previous studies on ICBT, although independent evaluations are necessary. We consider it a strength that the intervention was investigated with adolescents 15 to 19 years of age as opposed to a wider age range. It is indicated that the causes and constructs of depression may differ between children and adolescents [3], thus studies including both groups could possibly suffer from these differences interacting and interfering with treatment outcomes. A number of limitations should be considered: this study is one of the first on chat-supported ICBT, thus we believed that no evidence-based treatment control group would be appropriate. Nevertheless, it is a limitation that an active control intervention was not included. In line with discussions [21] regarding what is considered an intervention, we moreover labeled our control condition an attention control, as controls were thoroughly assessed and monitored and interacted with. The appropriateness of the label can be discussed. Treatment was open label and participants' awareness of their allocation may have affected self-reported outcomes; similarly, the clinical interviews were not blinded at posttreatment, which calls for caution when interpreting remission rates. The reduction on anxiety ratings cannot be explained by the treatment as the control group also improved. As the intervention focused on depression, it is possible that it was not effective enough to result in major reductions in anxiety in the treatment group. Participants in the study were almost entirely female, so results cannot be generalized to males. Depression is more prevalent among women but gender distributions for Web-based interventions and support can be even more skewed [87,95-96]. There are differences in internet use patterns and preferences between adolescent girls and boys [43], and possibly these differences interact with Web-based interventions as they currently are offered. It has been discussed that young boys are more likely than girls to seek help as a consequence of being influenced by others, and this could explain their relatively low enrollment in Web-based interventions, given that such enrollment is often more dependent on self-motivation and in many cases self-referral [87]. Our study recruited via social media, and moreover via an account focusing on coping with mental health issues; this may have influenced gender uptake as well as attracted particularly motivated participants. Although ours and previous findings point to the need of alterations in recruitment and design so that similar Web-based intervention can reach boys, the positive findings with girls should not be undervalued and the intervention should be regarded complementary among others. Not limiting to gender, focused (tailored) approaches rather than broad and universal interventions may be more relevant to desired target groups, and thus possibly more successful. For example, using a narrow age span, in our experience, facilitates the creation of content that is developmentally appropriate. Finally, while our positive findings may relate to the novel features of the intervention, the study contributes limited information on how the effects were achieved, for example, on the specific impact of therapist support. Lacking information on what factors influence outcomes in ICBT with youth is a common study limitation [34].

Future Research

The next step is to extend focus to uncovering the effect of therapist interaction, and other theoretical and contextual treatment components in ICBT, to contribute to our understanding of what components positively affect treatment outcomes in psychotherapy with youth depression and how they do so. Multiphase optimization strategies that include factorial

experimental designs pose as viable options [97]. Technology-based treatment enables new evaluation strategies and the current approach in particular produces vast amounts of data from patient-therapist correspondences that benefit from natural language processing to help understand and refine conversation [98,99]. Qualitative investigation will help to ensure that components perceived essential by young individuals are not lost but enhanced in further developments.

Conclusions

This study investigated treatment consisting of Web-based CBT self-help material and weekly therapist chat sessions, *blended treatment*, for adolescent depression. The intervention attracted adolescents in need of mental health assistance and demonstrated positive completion rates in combination with substantial improvement in depression symptoms. The results are similar to those in our previous study, further demonstrating the potential of a text-based blended model to deliver CBT in accordance with the urgent need for accessible behavioral intervention for youth.

Acknowledgments

NT was supported by Queen Silvia's Jubilee Fund, The Sweden-America Foundation, The Swedish Society of Medicine, and The Swedish Psychotherapy Society in this research; the study received funding from The Swedish Central Bank (P16-0883:1). The funders had no role in the conduct of the research and were not involved in review or approval of the manuscript for publication. NT authored the manuscript; all authors reviewed and approved the manuscript for publication. We thank George Vlaescu for excellent Web-master services, and Dr C Barr Taylor, Stanford University and Center for m^{2}Health, for comments on the manuscript. A special thanks to the adolescents who participated, and those who kindly distributed information about the study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-EHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 397 KB - jmir_v21i11e13393_app1.pdf]

References

1. Gore FM, Bloem PJ, Patton GC, Ferguson J, Joseph V, Coffey C, et al. Global burden of disease in young people aged 10-24 years: a systematic analysis. Lancet 2011 Jun 18;377(9783):2093-2102. [doi: 10.1016/S0140-6736(11)60512-6] [Medline: 21652063]
2. Kim-Cohen J, Caspi A, Moffitt TE, Harrington H, Milne BJ, Poulton R. Prior juvenile diagnoses in adults with mental disorder: developmental follow-back of a prospective-longitudinal cohort. Arch Gen Psychiatry 2003 Jul;60(7):709-717. [doi: 10.1001/archpsyc.60.7.709] [Medline: 12860775]
3. Thapar A, Collishaw S, Pine DS, Thapar AK. Depression in adolescence. Lancet 2012 Mar 17;379(9820):1056-1067 [FREE Full text] [doi: 10.1016/S0140-6736(11)60871-4] [Medline: 22305766]
4. Kessler RC. The costs of depression. Psychiatr Clin North Am 2012 Mar;35(1):1-14 [FREE Full text] [doi: 10.1016/j.psc.2011.11.005] [Medline: 22370487]
5. Lewinsohn P. Major depressive disorder in older adolescents. Prevalence, risk factors, and clinical implications. Clinical Psychology Review 1998 Nov;18(7):765-794. [doi: 10.1016/S0272-7358(98)00010-5]
6. Fletcher JM. Adolescent depression and educational attainment: results using sibling fixed effects. Health Econ 2010 Jul;19(7):855-871. [doi: 10.1002/hec.1526] [Medline: 19582699]
7. Fergusson DM, Horwood LJ, Ridder EM, Beautrais AL. Subthreshold depression in adolescence and mental health outcomes in adulthood. Arch Gen Psychiatry 2005 Jan;62(1):66-72. [doi: 10.1001/archpsyc.62.1.66] [Medline: 15630074]
8. Bardone AM, Moffitt TE, Caspi A, Dickson N, Stanton WR, Silva PA. Adult physical health outcomes of adolescent girls with conduct disorder, depression, and anxiety. J Am Acad Child Adolesc Psychiatry 1998 Jun;37(6):594-601. [doi: 10.1097/00004583-199806000-00009] [Medline: 9628079]
9. Copeland WE, Shanahan L, Costello EJ, Angold A. Childhood and adolescent psychiatric disorders as predictors of young adult disorders. Arch Gen Psychiatry 2009 Jul;66(7):764-772 [FREE Full text] [doi: 10.1001/archgenpsychiatry.2009.85] [Medline: 19581568]
10. Chisholm D, Sweeny K, Sheehan P, Rasmussen B, Smit F, Cuijpers P, et al. Scaling-up treatment of depression and anxiety: a global return on investment analysis. Lancet Psychiatry 2016 May;3(5):415-424 [FREE Full text] [doi: 10.1016/S2215-0366(16)30024-4]
11. Ebert DD, Zarski AC, Christensen H, Stikkelbroek Y, Cuijpers P, Berking M, et al. Internet and computer-based cognitive behavioral therapy for anxiety and depression in youth: a meta-analysis of randomized controlled outcome trials. PLoS One 2015;10(3):e0119895 [FREE Full text] [doi: 10.1371/journal.pone.0119895] [Medline: 25786025]

12. Butler AC, Chapman JE, Forman EM, Beck AT. The empirical status of cognitive-behavioral therapy: a review of meta-analyses. Clin Psychol Rev 2006 Jan;26(1):17-31. [doi: 10.1016/j.cpr.2005.07.003] [Medline: 16199119]
13. Cristea IA, Stefan S, Karyotaki E, David D, Hollon SD, Cuijpers P. The effects of cognitive behavioral therapy are not systematically falling: a revision of Johnsen and Friborg (2015). Psychol Bull 2017 Mar;143(3):326-340. [doi: 10.1037/bul0000062] [Medline: 28230413]
14. Driessen E, Hollon SD. Cognitive behavioral therapy for mood disorders: efficacy, moderators and mediators. Psychiatr Clin North Am 2010 Sep;33(3):537-555 [FREE Full text] [doi: 10.1016/j.psc.2010.04.005] [Medline: 20599132]
15. David D, Cristea I, Hofmann SG. Why cognitive behavioral therapy is the current gold standard of psychotherapy. Front Psychiatry 2018;9:4 [FREE Full text] [doi: 10.3389/fpsyt.2018.00004] [Medline: 29434552]
16. Wenzel A. Basic strategies of cognitive behavioral therapy. Psychiatr Clin North Am 2017 Dec;40(4):597-609. [doi: 10.1016/j.psc.2017.07.001] [Medline: 29080588]
17. Hofmann SG. An Introduction To Modern CBT: Psychological Solutions To Mental Health Problems. Oxford, UK: Wiley-Blackwell; 2011.
18. Andersson G, Titov N, Dear BF, Rozental A, Carlbring P. Internet-delivered psychological treatments: from innovation to implementation. World Psychiatry 2019 Feb;18(1):20-28 [FREE Full text] [doi: 10.1002/wps.20610] [Medline: 30600624]
19. Hedman E, Ljótsson B, Lindefors N. Cognitive behavior therapy via the internet: a systematic review of applications, clinical efficacy and cost-effectiveness. Expert Rev Pharmacoecon Outcomes Res 2012 Dec;12(6):745-764. [doi: 10.1586/erp.12.67] [Medline: 23252357]
20. Titov N, Dear BF, Staples LG, Terides MD, Karin E, Sheehan J, et al. Disorder-specific versus transdiagnostic and clinician-guided versus self-guided treatment for major depressive disorder and comorbid anxiety disorders: a randomized controlled trial. J Anxiety Disord 2015 Oct;35:88-102 [FREE Full text] [doi: 10.1016/j.janxdis.2015.08.002] [Medline: 26422822]
21. Baumeister H, Reichler L, Munzinger M, Lin J. The impact of guidance on internet-based mental health interventions — a systematic review. Internet Interv 2014 Oct;1(4):205-215. [doi: 10.1016/j.invent.2014.08.003]
22. Richards D, Richardson T. Computer-based psychological treatments for depression: a systematic review and meta-analysis. Clin Psychol Rev 2012 Jun;32(4):329-342. [doi: 10.1016/j.cpr.2012.02.004] [Medline: 22466510]
23. Johansson R, Andersson G. Internet-based psychological treatments for depression. Expert Rev Neurother 2012 Jul;12(7):861-9; quiz 870. [doi: 10.1586/ern.12.63] [Medline: 22853793]
24. Andersson G, Topooco N, Havik O, Nordgreen T. Internet-supported versus face-to-face cognitive behavior therapy for depression. Expert Rev Neurother 2016;16(1):55-60. [doi: 10.1586/14737175.2015.1125783] [Medline: 26610160]
25. Carlbring P, Andersson G, Cuijpers P, Riper H, Hedman-Lagerlöf E. Internet-based vs face-to-face cognitive behavior therapy for psychiatric and somatic disorders: an updated systematic review and meta-analysis. Cogn Behav Ther 2018 Jan;47(1):1-18. [doi: 10.1080/16506073.2017.1401115] [Medline: 29215315]
26. Ebert DD, Baumeister H. Internet-based self-help interventions for depression in routine care. JAMA Psychiatry 2017 Aug 1;74(8):852-853. [doi: 10.1001/jamapsychiatry.2017.1394] [Medline: 28636717]
27. Karyotaki E, Riper H, Twisk J, Hoogendoorn A, Kleiboer A, Mira A, et al. Efficacy of self-guided internet-based cognitive behavioral therapy in the treatment of depressive symptoms: a meta-analysis of individual participant data. JAMA Psychiatry 2017 Apr 1;74(4):351-359. [doi: 10.1001/jamapsychiatry.2017.0044] [Medline: 28241179]
28. Batterham PJ, Neil AL, Bennett K, Griffiths KM, Christensen H. Predictors of adherence among community users of a cognitive behavior therapy website. Patient Prefer Adherence 2008 Feb 2;2:97-105 [FREE Full text] [Medline: 19920949]
29. Holst A, Nejati S, Björkelund C, Eriksson MCM, Hange D, Kivi M, et al. Patients' experiences of a computerised self-help program for treating depression - a qualitative study of Internet mediated cognitive behavioural therapy in primary care. Scand J Prim Health Care 2017 Mar;35(1):46-53 [FREE Full text] [doi: 10.1080/02813432.2017.1288813] [Medline: 28277055]
30. Ye X, Bapuji SB, Winters SE, Struthers A, Raynard M, Metge C, et al. Effectiveness of internet-based interventions for children, youth, and young adults with anxiety and/or depression: a systematic review and meta-analysis. BMC Health Serv Res 2014 Jul 18;14(1):313-353 [FREE Full text] [doi: 10.1186/1472-6963-14-313] [Medline: 25037951]
31. Hollis C, Falconer CJ, Martin JL, Whittington C, Stockton S, Glazebrook C, et al. Annual Research Review: Digital health interventions for children and young people with mental health problems - a systematic and meta-review. J Child Psychol Psychiatry 2017 Apr 18;58(4):474-503. [doi: 10.1111/jcpp.12663] [Medline: 27943285]
32. Pennant ME, Loucas CE, Whittington C, Creswell C, Fonagy P, Fuggle P, Expert Advisory Group. Computerised therapies for anxiety and depression in children and young people: a systematic review and meta-analysis. Behav Res Ther 2015 Apr;67:1-18. [doi: 10.1016/j.brat.2015.01.009] [Medline: 25727678]
33. Grist R, Croker A, Denne M, Stallard P. Technology delivered interventions for depression and anxiety in children and adolescents: a systematic review and meta-analysis. Clin Child Fam Psychol Rev 2019 Jun;22(2):147-171 [FREE Full text] [doi: 10.1007/s10567-018-0271-8] [Medline: 30229343]
34. Wozney L, Huguet A, Bennett K, Radomski AD, Hartling L, Dyson M, et al. How do eHealth programs for adolescents with depression work? A realist review of persuasive system design components in internet-based psychological therapies. J Med Internet Res 2017 Aug 9;19(8):e266 [FREE Full text] [doi: 10.2196/jmir.7573] [Medline: 28793983]

35. Neil AL, Batterham P, Christensen H, Bennett K, Griffiths KM. Predictors of adherence by adolescents to a cognitive behavior therapy website in school and community-based settings. J Med Internet Res 2009 Feb 23;11(1):e6 [FREE Full text] [doi: 10.2196/jmir.1050] [Medline: 19275982]
36. Lillevoll KR, Vangberg HC, Griffiths KM, Waterloo K, Eisemann MR. Uptake and adherence of a self-directed internet-based mental health intervention with tailored e-mail reminders in senior high schools in Norway. BMC Psychiatry 2014 Jan 21;14:14 [FREE Full text] [doi: 10.1186/1471-244X-14-14] [Medline: 24443820]
37. Kleiboer A, Smit J, Bosmans J, Ruwaard J, Andersson G, Topooco N, et al. European comparative effectiveness research on blended depression treatment versus treatment-as-usual (E-COMPARED): study protocol for a randomized controlled, non-inferiority trial in eight European countries. Trials 2016 Aug 3;17(1):387 [FREE Full text] [doi: 10.1186/s13063-016-1511-1] [Medline: 27488181]
38. Topooco N. Blended Cognitive Behavior Therapy: Efficacy and Acceptability for Treating Depression in the Adult and Adolescent Population. Linköping: Linköping University Electronic Press; 2018.
39. Thase ME, Wright JH, Eells TD, Barrett MS, Wisniewski SR, Balasubramani G, et al. Improving the efficiency of psychotherapy for depression: computer-assisted versus standard CBT. Am J Psychiatry 2018 Mar 1;175(3):242-250 [FREE Full text] [doi: 10.1176/appi.ajp.2017.17010089] [Medline: 28969439]
40. Ly KH, Topooco N, Cederlund H, Wallin A, Bergström J, Molander O, et al. Smartphone-supported versus full behavioural activation for depression: a randomised controlled trial. PLoS One 2015;10(5):e0126559 [FREE Full text] [doi: 10.1371/journal.pone.0126559] [Medline: 26010890]
41. Topooco N, Riper H, Araya R, Berking M, Brunn M, Chevreul K, E-COMPARED Consortium. Attitudes towards digital treatment for depression: a European stakeholder survey. Internet Interv 2017 Jun;8:1-9 [FREE Full text] [doi: 10.1016/j.invent.2017.01.001] [Medline: 30135823]
42. Schuster R, Pokorny R, Berger T, Topooco N, Laireiter AR. The advantages and disadvantages of online and blended therapy: survey study amongst licensed psychotherapists in Austria. J Med Internet Res 2018 Dec 18;20(12):e11007 [FREE Full text] [doi: 10.2196/11007] [Medline: 30563817]
43. Lenhart A, Smith A, Anderson M, Duggan M, Perrin A. Pew Research Center. 2015. Teens, Technology and Friendships URL: https://www.pewinternet.org/2015/08/06/teens-technology-and-friendships/ [accessed 2019-02-01] [WebCite Cache ID 75ruDC47v]
44. Anderson M, Jiang J. Pew Research Center. 2018. Teens, Social Media & Technology 2018 URL: https://www.pewinternet.org/2018/05/31/teens-social-media-technology-2018/ [accessed 2019-09-18]
45. Rideout V, Fox S. Hopelab. 2018. Digital Health Practices, Social Media Use, and Mental Well-Being Among Teens and Young Adults in the US URL: https://www.hopelab.org/reports/pdf/a-national-survey-by-hopelab-and-well-being-trust-2018.pdf [accessed 2019-02-02] [WebCite Cache ID 75sEH61RA]
46. Davey CG, McGorry PD. Early intervention for depression in young people: a blind spot in mental health care. Lancet Psychiatry 2019 Mar;6(3):267-272. [doi: 10.1016/S2215-0366(18)30292-X]
47. Topooco N, Berg M, Johansson S, Liljethörn L, Radvogin E, Vlaescu G, et al. Chat- and internet-based cognitive-behavioural therapy in treatment of adolescent depression: randomised controlled trial. BJPsych Open 2018 Jul;4(4):199-207 [FREE Full text] [doi: 10.1192/bjo.2018.18] [Medline: 29988969]
48. Beck AT, Steer RA, Brown GK. Manual for the Beck Depression Inventory II. San Antonio, Texas: Psychological Corporation; 1996.
49. Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, et al. The Mini-International Neuropsychiatric Interview (MINI): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J Clin Psychiatry 1998;59(Suppl 20):22-33;quiz 34 [FREE Full text] [Medline: 9881538]
50. Vlaescu G, Alasjö A, Miloff A, Carlbring P, Andersson G. Features and functionality of the Iterapi platform for internet-based psychological treatment. Internet Interv 2016 Nov;6:107-114 [FREE Full text] [doi: 10.1016/j.invent.2016.09.006] [Medline: 30135819]
51. Dimidjian S, Barrera M, Martell C, Muñoz RF, Lewinsohn PM. The origins and current status of behavioral activation treatments for depression. Annu Rev Clin Psychol 2011;7:1-38. [doi: 10.1146/annurev-clinpsy-032210-104535] [Medline: 21275642]
52. Lewinsohn PM. A behavioral approach to depression. In: Friedman RJ, Katz MM, editors. The Psychology of Depression: Contemporary Theory and Research. New York: John Wiley & Sons Inc; 1974:157-185.
53. Hofmann SG, Smits JAJ. The evolution of cognitive behavioral therapy for anxiety and depression. Psychiatr Clin North Am 2017 Dec;40(4):xi-xii. [doi: 10.1016/j.psc.2017.08.011] [Medline: 29080599]
54. Daviss WB, Birmaher B, Melhem NA, Axelson DA, Michaels SM, Brent DA. Criterion validity of the mood and feelings questionnaire for depressive episodes in clinic and non-clinic subjects. J Child Psychol Psychiatry 2006 Sep;47(9):927-934. [doi: 10.1111/j.1469-7610.2006.01646.x] [Medline: 16930387]
55. Beck A, Epstein N, Brown G, Steer RA. An inventory for measuring clinical anxiety: psychometric properties. J Consult Clin Psychol 1988 Dec;56(6):893-897. [doi: 10.1037//0022-006x.56.6.893] [Medline: 3204199]
56. Mattick RP, Clarke JC. Development and validation of measures of social phobia scrutiny fear and social interaction anxiety. Behav Res Ther 1998 Apr;36(4):455-470. [doi: 10.1016/s0005-7967(97)10031-6] [Medline: 9670605]

57. Schwarzer R, Jerusalem M. Generalized self-efficacy scale. In: Weinman J, Wright S, Johnston M, editors. Measures in Health Psychology: A User's Portfolio. Windsor, UK: Nfer-Nelson; 1995:35-37.
58. Devilly GJ, Borkovec TD. Psychometric properties of the credibility/expectancy questionnaire. J Behav Ther Exp Psychiatry 2000 Jun;31(2):73-86. [doi: 10.1016/S0005-7916(00)00012-4] [Medline: 11132119]
59. Tracey TJ, Kokotovic AM. Factor structure of the working alliance inventory. Psychol Assess 1989 Sep;1(3):207-210. [doi: 10.1037/1040-3590.1.3.207]
60. Lindner P, Frykheden O, Forsström D, Andersson E, Ljótsson B, Hedman E, et al. The Brunnsviken brief quality of life scale (BBQ): development and psychometric evaluation. Cogn Behav Ther 2016 Apr;45(3):182-195 [FREE Full text] [doi: 10.1080/16506073.2016.1143526] [Medline: 26886248]
61. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. J Consult Clin Psychol 1991 Feb;59(1):12-19. [doi: 10.1037//0022-006x.59.1.12] [Medline: 2002127]
62. Evans C, Margison F, Barkham M. The contribution of reliable and clinically significant change methods to evidence-based mental health. Evid Based Ment Health 1998 Aug 1;1(3):70-72. [doi: 10.1136/ebmh.1.3.70]
63. Erbe D, Eichert HC, Riper H, Ebert DD. Blending face-to-face and internet-based interventions for the treatment of mental disorders in adults: systematic review. J Med Internet Res 2017 Sep 15;19(9):e306 [FREE Full text] [doi: 10.2196/jmir.6588] [Medline: 28916506]
64. Sethi S, Campbell AJ, Ellis LA. The use of computerized self-help packages to treat adolescent depression and anxiety. J Technol Human Serv 2010 Aug 31;28(3):144-160. [doi: 10.1080/15228835.2010.508317]
65. Sethi S. Treating youth depression and anxiety: a randomised controlled trial examining the efficacy of computerised versus face-to-face cognitive behaviour therapy. Aust Psychol 2013 Feb 14;48(4):249-257. [doi: 10.1111/ap.12006]
66. van der Zanden R, Kramer J, Gerrits R, Cuijpers P. Effectiveness of an online group course for depression in adolescents and young adults: a randomized trial. J Med Internet Res 2012 Jun 7;14(3):e86 [FREE Full text] [doi: 10.2196/jmir.2033] [Medline: 22677437]
67. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
68. Ip P, Chim D, Chan KL, Li TM, Ho FK, van Voorhees BW, et al. Effectiveness of a culturally attuned internet-based depression prevention program for Chinese adolescents: a randomized controlled trial. Depress Anxiety 2016 Dec;33(12):1123-1131. [doi: 10.1002/da.22554] [Medline: 27618799]
69. van Voorhees BW, Fogel J, Reinecke MA, Gladstone T, Stuart S, Gollan J, et al. Randomized clinical trial of an internet-based depression prevention program for adolescents (project CATCH-IT) in primary care: 12-week outcomes. J Dev Behav Pediatr 2009 Feb;30(1):23-37 [FREE Full text] [doi: 10.1097/DBP.0b013e3181966c2a] [Medline: 19194326]
70. Makarushka MM. Efficacy of an Internet-based Intervention Targeted to Adolescents with Subthreshold Depression. University of Oregon 2012 URL: https://scholarsbank.uoregon.edu/xmlui/handle/1794/12091 [accessed 2019-09-18]
71. Ellis LA, Campbell AJ, Sethi S, O'Dea BM. Comparative randomized trial of an online cognitive-behavioral therapy program and an online support group for depression and anxiety. J Cyber Ther Rehabil 2011;4(4):461-467 [FREE Full text]
72. Clarke G, Kelleher C, Hornbrook M, Debar L, Dickerson J, Gullion C. Randomized effectiveness trial of an internet, pure self-help, cognitive behavioral intervention for depressive symptoms in young adults. Cogn Behav Ther 2009;38(4):222-234 [FREE Full text] [doi: 10.1080/16506070802675353] [Medline: 19440896]
73. Smith P, Scott R, Eshkevari E, Jatta F, Leigh E, Harris V, et al. Computerised CBT for depressed adolescents: randomised controlled trial. Behav Res Ther 2015 Oct;73:104-110. [doi: 10.1016/j.brat.2015.07.009] [Medline: 26301756]
74. Merry SN, Stasiak K, Shepherd M, Frampton C, Fleming T, Lucassen MF. The effectiveness of SPARX, a computerised self help intervention for adolescents seeking help for depression: randomised controlled non-inferiority trial. Br Med J 2012 Apr 18;344:e2598 [FREE Full text] [doi: 10.1136/bmj.e2598] [Medline: 22517917]
75. Stasiak K, Hatcher S, Frampton C, Merry SN. A pilot double blind randomized placebo controlled trial of a prototype computer-based cognitive behavioural therapy program for adolescents with symptoms of depression. Behav Cogn Psychother 2014 Jul;42(4):385-401. [doi: 10.1017/S1352465812001087] [Medline: 23253641]
76. Fleming T, Dixon R, Frampton C, Merry S. A pragmatic randomized controlled trial of computerized CBT (SPARX) for symptoms of depression among adolescents excluded from mainstream education. Behav Cogn Psychother 2012 Oct;40(5):529-541. [doi: 10.1017/S1352465811000695] [Medline: 22137185]
77. Wright B, Tindall L, Littlewood E, Allgar V, Abeles P, Trépel D, et al. Computerised cognitive-behavioural therapy for depression in adolescents: feasibility results and 4-month outcomes of a UK randomised controlled trial. BMJ Open 2017 Jan 27;7(1):e012834 [FREE Full text] [doi: 10.1136/bmjopen-2016-012834] [Medline: 28132000]
78. Stallard P, Richardson T, Velleman S, Attwood M. Computerized CBT (Think, Feel, Do) for depression and anxiety in children and adolescents: outcomes and feedback from a pilot randomized controlled trial. Behav Cogn Psychother 2011 May;39(3):273-284. [doi: 10.1017/S135246581000086X] [Medline: 21272393]
79. Talkspace. Talkspace for Teens URL: https://lp.talkspace.com/try/talkspace-for-teens [accessed 2019-02-04]
80. Klein JB, Jacobs RH, Reinecke MA. Cognitive-behavioral therapy for adolescent depression: a meta-analytic investigation of changes in effect-size estimates. J Am Acad Child Adolesc Psychiatry 2007 Nov;46(11):1403-1413 [FREE Full text] [doi: 10.1097/chi.0b013e3180592aaa] [Medline: 18049290]

81. Weisz JR, McCarty CA, Valeri SM. Effects of psychotherapy for depression in children and adolescents: a meta-analysis. Psychol Bull 2006 Jan;132(1):132-149 [FREE Full text] [doi: 10.1037/0033-2909.132.1.132] [Medline: 16435960]
82. Yang L, Zhou X, Zhou C, Zhang Y, Pu J, Liu L, et al. Efficacy and acceptability of cognitive behavioral therapy for depression in children: a systematic review and meta-analysis. Acad Pediatr 2017;17(1):9-16. [doi: 10.1016/j.acap.2016.08.002] [Medline: 27989281]
83. Weisz JR, Kuppens S, Ng MY, Eckshtain D, Ugueto AM, Vaughn-Coaxum R, et al. What five decades of research tells us about the effects of youth psychological therapy: a multilevel meta-analysis and implications for science and practice. Am Psychol 2017;72(2):79-117. [doi: 10.1037/a0040360] [Medline: 28221063]
84. Vostanis P, Feehan C, Grattan E, Bickerton WL. Treatment for children and adolescents with depression: lessons from a controlled trial. Clin Child Psychol Psychiatry 2016 Jul 26;1(2):199-212. [doi: 10.1177/1359104596012003]
85. Weisz JR, Southam-Gerow MA, Gordis EB, Connor-Smith JK, Chu BC, Langer DA, et al. Cognitive-behavioral therapy versus usual clinical care for youth depression: an initial test of transportability to community clinics and clinicians. J Consult Clin Psychol 2009 Jun;77(3):383-396 [FREE Full text] [doi: 10.1037/a0013877] [Medline: 19485581]
86. Kennard B, Silva S, Vitiello B, Curry J, Kratochvil C, Simons A, TADS Team. Remission and residual symptoms after short-term treatment in the Treatment of Adolescents with Depression Study (TADS). J Am Acad Child Adolesc Psychiatry 2006 Dec 18;45(12):1404-1411. [doi: 10.1097/01.chi.0000242228.75516.21] [Medline: 17135985]
87. Rickwood D, Webb M, Kennedy V, Telford N. Who are the young people choosing web-based mental health support? Findings from the implementation of Australia's national web-based youth mental health service, eheadspace. JMIR Ment Health 2016 Aug 25;3(3):e40 [FREE Full text] [doi: 10.2196/mental.5988] [Medline: 27562729]
88. van Meter A, Cosgrove VE. Overhauling technology-based interventions for young people with bipolar disorder: lessons learned from adults. Bipolar Disord 2019 Feb;21(1):86-87. [doi: 10.1111/bdi.12716] [Medline: 30387928]
89. Torous J, Kossowsky J. Psychiatric Times. 2018. Digital Mental Health for Youth: New Evidence but Still Much Unknown URL: https://www.psychiatrictimes.com/telepsychiatry/digital-mental-health-youth-new-evidence-still-much-unknown [accessed 2019-01-22]
90. Warren JS, Nelson PL, Mondragon SA, Baldwin SA, Burlingame GM. Youth psychotherapy change trajectories and outcomes in usual care: community mental health versus managed care settings. J Consult Clin Psychol 2010 Apr;78(2):144-155. [doi: 10.1037/a0018544] [Medline: 20350026]
91. Karyotaki E, Kemmeren L, Riper H, Twisk J, Hoogendoorn A, Kleiboer A, et al. Is self-guided internet-based cognitive behavioural therapy (iCBT) harmful? An individual participant data meta-analysis. Psychol Med 2018 Nov;48(15):2456-2466 [FREE Full text] [doi: 10.1017/S0033291718000648] [Medline: 29540243]
92. Child and Adolescent Mental Health Services [BUP]. Stockholm County: Child and Adolescent Mental Health Services; 2016. Will parents always know? [Får föräldrar alltid veta?] URL: http://www.bup.se/sv/Rad-och-fakta/Artiklar/Bra-att-veta-om-BUP/Far-foraldrar-alltid-veta/
93. Santana B. Mental Health Practitioners' Understanding of Informed Consent with Adolescents. CUNY Academic Works: City University of New York; 2016 [FREE Full text] [accessed] 2019-09-18
94. National Center for Youth Law. 2010. New CA Minor Consent Law Increases Teens' Access to Mental Health Care URL: https://youthlaw.org/publication/new-ca-minor-consent-law-increases-teens-access-to-mental-health-care/ [accessed 2019-07-20] [WebCite Cache ID 75s2d5R5y]
95. Haner D, Pepler D. 'Live Chat' clients at kids help phone: individual characteristics and problem topics. J Can Acad Child Adolesc Psychiatry 2016;25(3):138-144 [FREE Full text] [Medline: 27924143]
96. Fukkink RG, Hermanns JM. Children's experiences with chat support and telephone support. J Child Psychol Psychiatry 2009 Jun;50(6):759-766. [doi: 10.1111/j.1469-7610.2008.02024.x] [Medline: 19207634]
97. Collins LM. Optimization Of Behavioral, Biobehavioral, And Biomedical Interventions: The Multiphase Optimization Strategy (MOST). New York: Springer; 2018.
98. Althoff T, Clark K, Leskovec J. Large-scale analysis of counseling conversations: an application of natural language processing to mental health. Trans Assoc Comput Linguist 2016;4:463-476 [FREE Full text] [doi: 10.1162/tacl_a_00111] [Medline: 28344978]
99. Hoogendoorn M, Funk B. Machine Learning for the Quantified Self: The Art of Learning from Sensory Data. Switzerland: Spinger International Publishing; 2017.

Abbreviations

ANCOVA: analysis of covariance
BAI: Beck Anxiety Inventory
BBQ: Brunnsviken Brief Quality of Life Scale
BDI-II: Beck Depression Inventory II
CBT: cognitive behavioral therapy
DSM-5: Diagnostic and Statistical Manual of Mental Disorder criteria
GSE: General Self-Efficacy Scale

ICBT: internet-delivered cognitive behavior therapy
ITT: intention-to-treat
MFQ: Mood and Feelings Questionnaire
MINI: Mini-International Neuropsychiatric Interview
SIAS: Social Interaction Anxiety Scale
WAI-S: Working Alliance Inventory

Edited by G Eysenbach; submitted 19.02.19; peer-reviewed by V Cosgrove, A Calear, S Rasing, D Rickwood; comments to author 18.04.19; revised version received 17.07.19; accepted 28.07.19; published 01.11.19.

Please cite as:
Topooco N, Byléhn S, Dahlström Nysäter E, Holmlund J, Lindegaard J, Johansson S, Åberg L, Bergman Nordgren L, Zetterqvist M, Andersson G
Evaluating the Efficacy of Internet-Delivered Cognitive Behavioral Therapy Blended With Synchronous Chat Sessions to Treat Adolescent Depression: Randomized Controlled Trial
J Med Internet Res 2019;21(11):e13393
URL: https://www.jmir.org/2019/11/e13393
doi:10.2196/13393
PMID:31682572

Original Paper

Use of a Fully Automated Internet-Based Cognitive Behavior Therapy Intervention in a Community Population of Adults With Depression Symptoms: Randomized Controlled Trial

Mark B Schure[1,2], MS, PhD; Janet C Lindow[2,3,4,5], PhD; John H Greist[2,3,6,7,8], MD; Paul A Nakonezny[9,10], PhD; Sandra J Bailey[1,2,11], PhD; William L Bryan[2,12], PhD; Matthew J Byerly[2,3,5,8], MD

[1]Department of Health & Human Development, Montana State University, Bozeman, MT, United States
[2]Center for Mental Health Research and Recovery, Montana State University, Bozeman, MT, United States
[3]Department of Psychiatry, College of Medicine, University of Arizona, Tucson, AZ, United States
[4]Biomedical Research and Education Foundation of Southern Arizona, Tucson, AZ, United States
[5]Southern Arizona VA Health Care System, Tucson, AZ, United States
[6]School of Medicine and Public Health, University of Wisconsin, Madison, WI, United States
[7]Healthcare Technology Systems, Madison, WI, United States
[8]Department of Cell Biology and Neuroscience, Montana State University, Bozeman, MT, United States
[9]Department of Population and Data Science, University of Texas Southwestern Medical Center, Dallas, TX, United States
[10]Department of Psychiatry, University of Texas Southwestern Medical Center, Dallas, TX, United States
[11]Montana State University Extension, Bozeman, MT, United States
[12]One Montana, Bozeman, MT, United States

Corresponding Author:
Mark B Schure, MS, PhD
Department of Health & Human Development
Montana State University
305 Herrick Hall
Bozeman, MT
United States
Phone: 1 406 994 3248
Email: mark.schure@montana.edu

Abstract

Background: Although internet-based cognitive behavior therapy (iCBT) interventions can reduce depression symptoms, large differences in their effectiveness exist.

Objective: The aim of this study was to evaluate the effectiveness of an iCBT intervention called Thrive, which was designed to enhance engagement when delivered as a fully automated, stand-alone intervention to a rural community population of adults with depression symptoms.

Methods: Using no diagnostic or treatment exclusions, 343 adults with depression symptoms were recruited from communities using an open-access website and randomized 1:1 to the Thrive intervention group or the control group. Using self-reports, participants were evaluated at baseline and 4 and 8 weeks for the primary outcome of depression symptom severity and secondary outcome measures of anxiety symptoms, work and social adjustment, psychological resilience, and suicidal ideation.

Results: Over the 8-week follow-up period, the intervention group (n=181) had significantly lower depression symptom severity than the control group (n=162; P<.001), with a moderate treatment effect size (d=0.63). Moderate to near-moderate effect sizes favoring the intervention group were observed for anxiety symptoms (P<.001; d=0.47), work/social functioning (P<.001; d=0.39), and resilience (P<.001; d=0.55). Although not significant, the intervention group was 45% less likely than the control group to experience increased suicidal ideation (odds ratio 0.55).

Conclusions: These findings suggest that the Thrive intervention was effective in reducing depression and anxiety symptom severity and improving functioning and resilience among a mostly rural community population of US adults. The effect sizes associated with Thrive were generally larger than those of other iCBT interventions delivered as a fully automated, stand-alone intervention.

Trial Registration: ClinicalTrials.gov NCT03244878; https://clinicaltrials.gov/ct2/show/NCT03244878

(J Med Internet Res 2019;21(11):e14754) doi:10.2196/14754

KEYWORDS

internet-based cognitive behavior therapy; iCBT; depression symptoms; rural populations; RCT; randomized controlled trial; CBT

Introduction

Depression is the leading cause of disability globally [1] and is associated with impaired function [2], higher morbidity and mortality [3,4], greater health care use [5], and higher risk of suicide [6]. In the United States, the 12-month and lifetime prevalence of major depressive disorder is 10.4% and 20.6%, respectively [2].

Cognitive behavior therapy (CBT), when delivered by clinicians, is an evidence-based psychotherapy for treating acute depression symptoms that also reduces relapse risk [7-9]. However, many barriers to receiving CBT exist, including geographic location, cost of insurance and care, long waitlists, and stigma for seeking treatment for mental illness [10]. For the approximately 46 million Americans living in rural regions [11], access to adequate mental health care is often limited or nonexistent [12]. In Montana (the focus of this study), 65.0% of residents live in rural regions [13]; the state also has a higher prevalence of depression than the national average and has had one of the highest suicide rates in the United States over the past four decades [14,15]. Despite the state's clear need for mental health services, 60% of the population live in mental health care professional shortage areas (defined as <1 psychiatrist/30,000) and only 13.3% of the total mental health provider need is met [12].

An attractive option for treating mental disorders in rural regions is the use of affordable, internet-delivered psychotherapy that requires little or no human intervention. Several studies of traditional face-to-face CBT and internet-based CBT (iCBT), herein defined as a class of online software apps that emulate in-person psychotherapy, have shown similar effectiveness for treating depression symptoms [16-18]. Additionally, a recent meta-analysis showed comparable effectiveness of iCBT interventions for reducing depression and anxiety symptoms among urban and rural populations outside the United States [19]. These results, combined with increasing broadband access in US rural communities [20], suggest that effective iCBT interventions have the potential for widespread public health impact by expanding the availability of low-cost, effective depression treatments [21,22] and providing an attractive alternative or complementary delivery strategy for populations that face the aforementioned barriers. However, no randomized controlled trials (RCTs) of iCBT interventions for depression have been conducted among US adult rural residents [21,22].

Although, in theory, the use of iCBT interventions to treat depression symptoms in US regions lacking sufficient mental health care services is an attractive treatment strategy, determining which iCBT intervention might be most effective is complicated by the considerable differences that exist in iCBT program design (eg, static text and pictures vs video-centric formats), support (fully automated to extensive human supports), adherence, and demonstrated effectiveness for treating depression symptoms in adults with a range of symptom severities (mild to severe) and diagnoses (none, major depressive disorder, unipolar affective disorder, or dysthymia) [10,21,22]. Additionally, participant engagement has been a significant challenge for some depression iCBT interventions. Fully automated, stand-alone iCBT interventions generally have higher treatment dropout rates (74%) than those provided with therapist or administrative support (28% and 38%, respectively) [23]. Lower engagement likely decreases the effectiveness of iCBT interventions for depression. For example, in three studies that reported participants completing fewer than three mean intervention sessions, two found no significant difference in depression symptoms for the intervention compared to the control [24,25] and the third found a small effect size (0.26) [26] for depression symptoms. To date, six of the seven studies of fully automated, stand-alone versions of iCBT have demonstrated small-to-medium clinical effects [26-31] and one found no significant effects [25]. Thus, improvements in iCBT interventions and delivery strategies are needed.

The intervention evaluated in this study, Thrive, is a fully automated, stand-alone iCBT intervention designed to reduce depression symptoms using a video-based platform. The intervention incorporates classic cognitive behavior therapy themes in modules on Constructive Thinking (Cognitive Restructuring) [32], Pleasant Activities (Behavioral Activation) [33,34], and Assertive Communication (Social Skills Training) [35]. Each module consists of 10 lessons, and there is an introduction lesson that offers suggestions for choice of a first module (31 total lessons). Within each module, Thrive's algorithms personalize content, exercises, and recommendations for participants based on their input and progress. A pilot feasibility study of Thrive in a US primary health care setting included 37 patients with depression (Symptom Checklist-20 score>1.75), of whom 59% (22/37) had suicidal thoughts at baseline (Symptom Checklist-20 item 13) [36]. At the 4-month follow-up, 52% (16/31) had ≥50% reduction in depression, 46% (14/31) had a clinically significant decrease in depression symptoms, and fewer reported having recent suicidal thoughts (35% vs 59% at baseline) [36]. Thrive was also offered to employees of four businesses located in four states. A total of 227 individuals with baseline depression symptom severity (Patient Health Questionnaire-9 item [PHQ-9] [37]) scores between 5 and 27 (mean score 10.5) and at least one follow-up PHQ-9 score logged into Thrive an average of 10.5 times and experienced an average PHQ-9 improvement of 4.4 points (42%) over an average of 7 weeks (B. Coleman, personal communication, 3 Sept 2019). Thus, preliminary data suggested that Thrive might be efficacious. Given the lack of evidence of

iCBT depression interventions in rural US settings, the primary objective of this RCT was to evaluate the short-term effectiveness of Thrive to reduce depression symptom severity when delivered directly to a community population of adults with depression symptoms living in Montana, one of the least densely populated states in the United States [38]. Because the Thrive intervention required no clinician interaction or support for delivery and assessments, findings are likely predictive of the use and impact of the Thrive intervention in real-world settings and support the promise of iCBT to address unmet mental health care needs in rural US regions and possibly beyond.

Methods

Trial Design

An RCT compared the efficacy of the fully automated, stand-alone Thrive intervention to a waitlist control (WLC) in reducing depression symptoms among adults. Participants were recruited from communities across Montana and immediately randomized 1:1 to the intervention group (Thrive) or a WLC group (delayed access to Thrive until the 8-week follow-up assessment) after meeting the inclusion criteria and providing electronic informed consent on the study website [39] (Multimedia Appendix 1). Participants in both groups were assessed for primary and secondary outcomes at baseline and 4 and 8 weeks postenrollment. Enrollment occurred between September 2017 and January 2018, and all assessments were completed by March 2018. The Montana State University Institutional Review Board approved the protocol and all related materials (#MS033017-FC) prior to study initiation. The study is registered at ClinicalTrials.gov (NCT03244878).

Participants

Participants were recruited using community fliers, public service announcements, local newspaper advertisements, newsletters, and community social media sites. The study was also promoted through select state organization email listservs, large employers, local health care providers, Facebook, a Craigslist community page, community meeting events, and Montana State University Extension faculty communications with their respective counties. All methods directed potential participants to a study website, which informed potential applicants about the study expectations; determined their eligibility; and guided those eligible through the informed consent, randomization, and assessment process [39].

Inclusion criteria included adults aged ≥18 years with mild-to-severe depression severity (PHQ-9 score >5) [37]; Montana residency; a valid email address; and regular access to broadband internet via a computer, tablet, or smartphone. At enrollment, potential participants who indicated recent suicidal ideation (PHQ-9 item 9 score >0) were asked to confirm that they could stay safe and those responding that they could not were considered ineligible (see the safety protocol description below for the handling of cases of suicidal ideation and Multimedia Appendix 4). All participants provided electronic informed consent prior to study participation. All participants were informed that they were free to obtain and use any additional care available throughout their participation in the trial. In total, 463 participants were enrolled, and of these, 109 were deemed fraudulent identities and their data were removed prior to analyses (Multimedia Appendix 2). Additionally, three individuals provided invalid email addresses, which led to their exclusion from study participation, and data from two control participants were discarded, as they were accidentally provided access to the intervention immediately. Six participants were excluded because of missing baseline data on the covariate "currently receiving psychosocial therapy for depression," which was required for data analysis. The final analytic sample included 343 participants (intervention: n=181; control: n=162), of which 86 (25%) participants did not complete any follow-up assessments (Multimedia Appendices 2 and 3).

Intervention Group

Thrive is a fully automated, stand-alone, individually tailored iCBT intervention for depression developed by Waypoint Health Innovations [40]. Participants accessed the intervention with a Web browser or mobile app. Intervention content is largely delivered by video with minimal text and employs three structured interactive modules focused on behavioral activation, cognitive restructuring, and social skills training CBT techniques. Based on user input and usage patterns, the intervention uses algorithms to personalize feedback and tailor user progression through each therapeutic modality. Based on the study results from a qualitative study on the acceptability of Thrive [41], 6 of the 320 videos were replaced with videos depicting scenarios and settings characteristic of rural Montana to enhance engagement among the study population.

Control Group

The control group received a link to general depression information at the National of Institute of Mental Health [42]. Participants in the control group were also provided a link to Montana's National Alliance on Mental Illness resource page [43]. The control group was granted access to Thrive after completing the 8-week assessment.

Assessments

Participants completed study assessments on the study website at baseline, 4 weeks, and 8 weeks. Automated email reminders were sent when each assessment was due, with two additional follow-up automated reminder emails sent within 7 days to those who had not completed the assessment. If the participant did not complete the scheduled assessment within a 10-day window, data were considered lost to follow-up for that assessment (Multimedia Appendix 10). A US $25 Amazon gift code was sent following completion of each assessment.

Outcomes

The primary outcome measure was participants' self-reported depression symptom severity PHQ-9 score. Prespecified secondary outcomes included anxiety symptom severity, daily functioning, and resilience. Suicidal ideation was added as a secondary outcome after the original protocol was developed, but prior to the start of analyses.

Measures

All outcomes were assessed using self-reported, validated measures automatically administered via the study website.

Depression symptom severity was evaluated using the PHQ-9, which incorporates the major depressive disorder "A criterion" symptoms of Diagnostic and Statistical Manual of Mental Disorders (DSM)-IV and DSM-5 (score range: 0-27, with higher scores indicating worse symptoms) [37]. Anxiety symptom severity was measured with the Generalized Anxiety Disorder Scale (GAD-7, score range: 0-21), with higher scores indicating worse symptoms [44]. Daily functioning was measured using the Work and Social Adjustment Scale (WSAS, score range: 0-40), with higher scores indicating a greater impact of depression on daily functioning [45]. The abbreviated version of the Conner-Davidson Resilience Scale (CD-RISC-10) measured resilience (score range: 0-40), with higher scores indicating greater resilience [46]. Lastly, suicidal ideation was assessed using item 9 of the PHQ-9 measure ("Thoughts that you would be better off dead or of hurting yourself" in the past two weeks). Item 9 was treated as an ordinal scale that ranged from 0 ("not at all," no suicidal ideation) to 3 ("nearly every day") [37].

Safety

When study applicants endorsed having at least some suicidal ideation (PHQ-9 item 9 score>0) at study enrollment, the study website displayed multiple sources of immediate help (Multimedia Appendix 4) and asked applicants to declare whether they were sure they could stay safe. Applicants answering that question negatively were considered ineligible to participate in the study. Notably, no applicants reported that they could not stay safe. Regardless of the response to the "stay safe" question, the study website prompted the individual to seek help and provided the same sources of help described in Multimedia Appendix 4. If an enrolled participant indicated s/he had at least some suicidal ideation during the 4- or 8-week assessments (PHQ-9 item 9 score >0), the study website immediately prompted the individual to seek help and provided the same sources of help. Additionally, the Thrive intervention directed individuals with PHQ-9 scores ≥20 at any assessment and those with PHQ-9 scores ≥10 on the third assessment to seek help from a doctor. A thorough description of all safety measured used is described in Multimedia Appendix 4. Participants were provided email addresses (lead investigator and IRB Chair) for reporting of any adverse experiences or events.

Covariates

An initial pool of 10 variables was selected *a priori* for analysis as potential covariates of depression severity, anxiety, functional impairment, resilience, and suicidal ideation (Multimedia Appendix 5).

Statistical Analysis

Outcomes were assessed at baseline and at 4 and 8 weeks. The change over time in each continuous outcome and suicidal ideation was compared between the intervention and control groups using a linear mixed model analysis of repeated measures and an ordinal logistic regression model within a Generalized Estimating Equation framework, respectively. A separate model was conducted on each outcome measure. Each model contained fixed-effects terms for treatment (intervention vs control), time, treatment × time interaction, and respective baseline measure (prior to the intervention) as covariates. Receiving therapy for depression (yes/no) at the baseline assessment was also included as a covariate in each model. Least squares means (adjusted treatment means) and adjusted odds ratios (OR) were estimated as part of the mixed model and ordinal logistic model, respectively, to interpret the treatment effect. For the ordinal logistic regression, the cumulative probabilities were modeled over the higher-ordered suicidal ideation scale scores (more suicidal ideation).

Statistical analyses were performed using SAS software, version 9.4 (SAS Institute, Inc, Cary, North Carolina). Maximum likelihood estimators allow efficient parameter estimation using only available data under an assumption of missing at random [47-49]. The level of significance was set at α=0.05 (two-tailed), and the Bonferroni method was implemented to control false-positives over the multiple tests. The 95% CIs for the point estimates of treatment group effects were also adjusted to match the Bonferroni adjustment to significance levels in the corresponding test. A priori evaluable sample size for a statistical power of 80% was estimated (Multimedia Appendix 6).

Results

Participant Characteristics

Participant baseline characteristics are described in Table 1. The cohort was predominately female (290/343, 85%) and Caucasian (319/343, 93%), with a mean age of 42.9 (SD 13.3) years. Most participants were married (198/343, 58%); employed, or in school (265/343, 77%); had some higher education (318/343, 93%); and lived in rural regions (287/343, 84%). In addition, 59% (202/343) of participants were receiving medication, psychosocial therapy, or other treatment for depression or anxiety at baseline. No significant differences for any outcome measure were observed at baseline (Table 2) nor did the baseline depression symptom severity differ between groups (Multimedia Appendix 7). Participants randomized to the intervention group completed a mean of 8.7 lessons. No adverse experiences or adverse events were reported by either study group.

Table 1. Baseline characteristics in the intent-to-treat population.

Characteristics	Total sample (N=343), n (%)	iCBT[a] group[b] (n=181), n (%)	Control group (n=162), n (%)
Age, mean (SD)	42.9 (SD 13.3)	42.1 (12.8)	43.8 (13.8)
Female	290 (85.0)	160 (88.4)	130 (81.2)
Race			
White	319 (93.0)	170 (93.9)	149 (92.0)
Other	24 (7.0)	11 (6.1)	13 (8.0)
Marital status			
Single	145 (42.3)	78 (43.1)	67 (41.4)
Married/domestic partnership	198 (57.7)	103 (56.9)	95 (58.6)
Employment status			
Employed or student	265 (77.3)	136 (75.1)	129 (79.6)
Retired or unemployed	78 (22.7)	45 (24.9)	33 (20.4)
Veteran	15 (4.4)	7 (3.9)	8 (4.9)
Education			
High school degree or less	25 (7.3)	10 (5.5)	15 (9.3)
Some college, bachelor or associate degree, or trade school	236 (68.8)	129 (71.3)	107 (66.0)
Graduate or professional degree	82 (23.9)	42 (23.2)	40 (24.7)
Health insurance			
Private	244 (71.1)	140 (77.3)	104 (64.2)
Public	74 (21.6)	32 (17.7)	42 (25.9)
Other	9 (2.6)	5 (2.8)	4 (2.5)
None	16 (4.7)	4 (2.2)	12 (7.4)
Rural classification[c]			
Urban	56 (16.3)	27 (14.9)	29 (17.9)
Large rural	73 (21.3)	42 (23.2)	31 (19.1)
Small rural	116 (33.8)	60 (33.2)	56 (34.6)
Isolated	98 (28.6)	52 (28.7)	46 (28.4)
Receiving mental health treatment[d]			
Yes	202 (58.9)	103 (56.9)	99 (61.1)
No	141 (41.1)	78 (43.1)	63 (38.9)

[a]iCBT: internet-based cognitive behavior therapy.

[b]Thrive intervention group.

[c]Defined using rural urban commuting area codes [50].

[d]Defined as receiving any clinical care or taking medication(s) for depression symptoms.

Table 2. Baseline clinical measures for total sample and treatment groups.

Measure	Total sample (N=343), mean (SD)	iCBT[a] group[b] (n=181), mean (SD)	Control group (n=162), mean (SD)
Primary outcome measure			
Depression symptom severity[c]	13.6 (5.0)	13.7 (5.0)	13.4 (5.0)
Secondary outcome measures			
Anxiety symptom severity[d]	10.3 (4.7)	10.3 (4.7)	10.2 (4.6)
Work and social functioning[e]	20.2 (7.9)	20.2 (8.0)	20.2 (7.8)
Resilience[f]	22.2 (6.2)	22.2 (6.6)	22.2 (5.9)

[a]iCBT: internet-based cognitive behavior therapy.

[b]Thrive intervention group.

[c]Patient Health Questionnaire-9 score range=0-27.

[d]Generalized Anxiety Disorder Scale 7-Item score range=0-21.

[e]Work and Social Adjustment Scale score range=0-40.

[f]Connor-Davidson Resilience Scale 10-Item score range=0-40.

Clinical Outcomes

Primary Outcome

Significant main effects of treatment ($F_{1,248}$=28.67; raw P<.001, adjusted P<.001) and time ($F_{1,216}$=11.94; raw P<.001, adjusted P<.001), favoring the Thrive intervention, were observed for depression symptom severity (Table 3), and a moderate treatment effect size (Cohen d=0.63, P<.001; Table 4) was found. The pattern of the overall least squares treatment group means showed that depression severity (following 8 weeks of intervention) was significantly lower for the intervention group than for the control group (7.702 [SE 0.336] vs 10.224 [SE 0.328], raw P<.001, adjusted P<.001; d=0.63; Table 4). The same pattern was observed with simple treatment group effects at weeks 4 and 8 (raw P<.001, adjusted P<.001; PHQ-9, Table 4).

Table 3. Main effects of treatment, time, and treatment by time interaction effects from the mixed model and ordinal logistic regression analysis for depression symptom severity, anxiety symptom severity, work/social functioning, resilience, and suicidal ideation.

Covariates and effects	Depression symptoms (PHQ-9[a])		Anxiety symptoms (GAD-7[b])		Functioning (WSAS[c])		Resilience (CD-RISC-10[d])		Suicidal ideation (PHQ-9)	
	F statistic (*df*)	*P* value	*F* statistic (*df*)	*P* value	*F* statistic (*df*)	*P* value	*F* statistic (*df*)	*P* value	χ^2 (*df*)	*P* value
Baseline outcome[e]	F (1, 245) =119.76	<.001	F (1, 253.6) =133.56	<.001	F (1, 251.7) =181.17	<.001	F (1, 247.7) =323.87	<.001	67.18 (1)	<.001
Therapy[f]	F (1, 249) =7.58	.006	F (1, 254.1) =8.29	.004	F (1, 250.7) =8.82	.003	F (1, 243.7) =4.17	.04	3.71 (1)	.054
Effects										
Treatment[g]	F (1, 248) =28.67	<.001; <.001[h]	F (1, 252.8) =16.14	<.001; <.001[h]	F (1, 249.4) =11.09	.001; .005[h]	F (1, 243.6) =22.71	<.001; <.001[h]	2.74 (1)	.098; .49[h]
Time (weeks)	F (1, 216) =11.94	<.001; <.001[h]	F (1, 221) =4.68	.03; .16[h]	F (1, 217.5) =0.49	.49; >.99[h]	F (1, 213.3) =5.82	.02; .08[h]	5.19 (1)	.02; .11[h]
Treatment × time	F (1, 216) =0.12	.73; >.99[h]	F (1, 221) =0.43	.51; >.99[h]	F (1, 217.5) =3.66	.06; .29[h]	F (1, 213.3) =0.22	.64; >.99[h]	1.39 (1)	.24; >.99[h]

[a]PHQ-9: Patient Health Questionnaire-9.

[b]GAD-7: Generalized Anxiety Disorder 7-Item scale.

[c]WSAS: Work and Social Adjustment Scale.

[d]CD-RISC-10: Connor-Davidson Resilience Scale 10-Item

[e]Baseline scores for PHQ-9, GAD-7, WSAS, and CD-RISC-10.

[f]Receiving treatment for depression at baseline.

[g]Thrive iCBT intervention or control.

[h]*P* values adjusted by the Bonferroni method.

Table 4. Effect of the intervention on depression severity, anxiety severity, work/social functioning, and resilience.

Outcome and group	Week 4, LSM[a] (SE), (95% CI)	Week 8, LSM (SE), (95% CI)	Overall timed-average (weeks 4-8), LSM (SE), (95% CI)	Overall treatment group main effect (weeks 4-8)		
				F statistic (*df*)	*P* value	Cohen *d*
Depression severity (PHQ-9[b])						
Intervention group	8.165 (0.373), (7.430 to 8.899)	7.240 (0.385), (6.481 to 7.998)	7.702 (0.336), (7.039 to 8.365)	N/A[c]	N/A	N/A
Control group	10.602 (0.371), (9.873 to 11.33)	9.845 (0.367), (9.122 to 10.568)	10.224 (0.328), (9.576 to 10.871)	N/A	N/A	N/A
LSM group difference	N/A	N/A	–2.521 (0.471), (–3.448 to –1.593), (–3.743 to –1.298)[d]	F (1, 248)=28.67	<.001; <.001[e]	0.628
Anxiety severity (GAD-7[f])						
Intervention group	7.121 (0.346), (6.439 to 7.803)	6.481 (0.358), (5.777 to 7.185)	6.801 (0.312), (6.186 to 7.415)	N/A	N/A	N/A
Control group	8.724 (0.344), (8.047 to 9.401)	8.382 (0.341), (7.710 to 9.053)	8.553 (0.304), (7.953 to 9.153)	N/A	N/A	N/A
LSM group difference	N/A	N/A	–1.752 (0.436), (–2.610 to –0.893), (–2.883 to –0.620)[d]	F (1, 252.8)=16.14	<.001; <.001[e]	0.470
Work/social functioning (WSAS[g])						
Intervention group	16.433 (0.602), (15.249 to 17.617)	15.407 (0.621), (14.185 to 16.629)	15.920 (0.542), (14.852 to 16.988)	N/A	N/A	N/A
Control group	18.205 (0.597), (17.029 to 19.380)	18.682 (0.592), (17.517 to 19.848)	18.443 (0.529), (17.401 to 19.486)	N/A	N/A	N/A
LSM group difference	N/A	N/A	–2.523 (0.757), (–4.016 to –1.031), (–4.487 to –0.558)[d]	F (1, 249.4)=11.09	.001; <.005[e]	0.389
Resilience (CD-RISC-10[h])						
Intervention group	24.899 (0.382), (24.148 to 25.652)	25.646 (0.395), (24.868 to 26.424)	25.273 (0.341), (24.601 to 25.945)	N/A	N/A	N/A[c]
Control group	22.749 (0.379), (22.003 to 23.496)	23.254 (0.376), (22.514 to 23.995)	23.002 (0.332), (22.346 to 23.657)	N/A	N/A	N/A[c]
LSM group difference	N/A	N/A	2.271 (0.476), (1.332 to 3.209), (1.035 to –3.506)[d]	F (1, 243.6)=22.71	<.001; <.001[e]	0.552

[a]LSM: least squares means.

[b]PHQ-9: Patient Health Questionnaire-9.

[c]N/A: not applicable.

[d]Bonferroni-adjusted 95% CIs.

[e]*P* values adjusted by the Bonferroni method. The adjusted *P* value was associated with the test (F statistic) of the overall timed-average difference of the LSM estimate between the groups (Thrive intervention vs control).

[f]GAD-7: Generalized Anxiety Disorder 7-Item scale.

[g]WSAS: Work and Social Adjustment Scale.

[h]CD-RISC-10: Connor-Davidson Resilience Scale 10-Item.

Secondary Outcomes

The results of the main effects (treatment and time) as well as the treatment by time interaction effect from the mixed model and ordinal logistic regression analysis for the secondary outcomes of anxiety, functional impairment, resilience and suicidal ideation are reported in Table 3. Anxiety symptom severity, work and social functional impairment, and resilience were significantly improved for the Thrive intervention group compared to the control group (raw $P<.001$, adjusted $P<.001$ for all; $d=0.47$, $d=0.39$, and $d=0.55$, respectively; Table 4). This pattern was also observed with simple treatment group effects at weeks 4 and 8 (Multimedia Appendix 8).

Suicidal ideation was reported by 41%, 19%, and 16% of participants at baseline, week 4, and week 8, respectively. The predicted odds of increased suicidal ideation (PHQ-9 item 9) for the intervention group showed a lower trend than that of the control group, but did not reach significance (OR 0.55, 95% CI 0.26-1.11, $P=.10$, adjusted $P=.49$; Multimedia Appendix 9). Thrive intervention group participants were 45% and 58% less likely than controls to experience increased suicidal ideation following the entire 8-week follow-up period and at week 8, respectively (Multimedia Appendix 9).

Discussion

Principal Results

This study evaluated the effectiveness of a fully automated, stand-alone, video-centric iCBT intervention (Thrive) for reducing depression symptom severity among adults residing in Montana. The design incorporated features of practical, pragmatic, and community-based effectiveness trials: liberal inclusion criteria (mild to severe depression symptoms), allowance for any past and concurrent treatments, and broad (symptom, functional, and resilience) outcomes. Over 8 weeks, depression symptom severity (primary outcome) in the Thrive intervention group was significantly reduced as compared to the control group. Additionally, the Thrive intervention group showed significant improvements in anxiety symptoms, work and social functioning, and resilience compared with the control group.

Although the Thrive intervention, like other iCBT interventions for depression in non-US populations [19], may be equally effective in rural and urban populations, the intervention has several characteristics that may make it appealing to rural US residents as an alternative or supplemental mental health intervention for depression symptoms. A previous qualitative study of Thrive among rural Montana adults led to an adaptation of 6 of the >300 videos to better reflect scenarios and settings common in rural communities [41], which may improve user engagement. Thrive can be performed in the privacy of one's home, which might offset mental health-related stigma, a frequent barrier to seeking mental health services experienced by rural US residents [51]. Thrive is accessible from any location with internet access and can be used on smartphones, tablets, and computers, making it an easily accessible, cost-effective alternative for rural residents who often have to drive long distances for in-person mental health care [41].

Comparison With Prior Work

To our knowledge, the clinical effect of the Thrive intervention on depression symptoms is one of the largest reported for an iCBT intervention delivered without clinician interaction. Omitting clinician support allows for broader generalizability to community populations by limiting costs and avoiding barriers associated with an underresourced mental health workforce. A recent meta-analysis of 16 self-guided iCBT studies by Karyotaki et al [21] and a systematic review by Lorenzo-Luaces et al [22] of the same 16 and an additional 5 self-guided iCBT studies reported that 7 studies evaluated iCBT interventions that, like Thrive, had no clinician contact [25-31]. The impact of Thrive on depression symptoms ($d=0.63$) was greater than or similar to the 7 comparable studies (nonsignificant [25]; $d=0.17$ [26]; $d=0.20$ [28]; $d=0.28$ [27]; $d=0.30$ and 0.65 [for analysis of variance and mixed model analyses, respectively] [29]; $g=0.36$ [31]; and $d=0.50$ [30]). The greater impact of Thrive may be a result of the intervention group completing more lessons on average (8.7) than some other fully automated, stand-alone iCBT interventions [24-26].

Compared to self-guided iCBT interventions delivered with meaningful clinician support, described in Karyotaki et al [21] and Lorenzo-Luaces et al [22], the Thrive intervention yielded a similar or greater effect size for depression symptoms: Four studies that used clinician/research personnel contact to determine eligibility found effect sizes of $d=0.08$ [52], $d=0.38$ [53], $d=0.8$ [54], and $g=0.76$ for depression symptoms [55], although the latter study excluded "outliers" from the primary analysis and stated their inclusion resulted in "nonsignificant findings" and reduced the effect size by an undisclosed amount [55]. Two other studies that used iCBT in combination with contact with a therapist also reported similar or lower effect sizes ($d=0.51$ [56] and 0.38 [31]) than those reported for Thrive. Six studies that included individualized email or SMS check-in reminders or telephone diagnostic interviews produced nonsignificant [57] or small-to-medium effect sizes for depression symptoms ($d=0.34$ [30]; $d=0.36$ [58]; $d=0.55$ [59]; $d=0.57$ [60]; and $d=0.66$ [61]). Finally, the six studies providing up to weekly nonautomated email or telephone support reported no significant differences compared to the control for two studies [24,62] and small-to-large significant effects sizes for four studies: (1) $d=0.22$ and $d=0.34$ for support upon request and weekly support, respectively [26]; (2) $d=0.4$ (pre-post within-subjects effect size) [63]; (3) $d=0.39$ [64]; and (4) $d=1.14$ [61].

This study found similar or greater effect sizes for secondary outcomes (anxiety symptoms [$d=0.47$], work and social functioning [$d=0.39$], and resilience [$d=0.55$]) compared to the 21 iCBT trials described in the recent meta-analysis and systematic review [21,22]. In the seven studies reporting anxiety symptom outcomes, four identified nonsignificant differences with the control [53,56,60,62] and three found small-to-nearly moderate effect sizes ($d=0.22$ [54], $d=0.25$ [26], and $d=0.43$ [30]). The three studies that assessed life/work functioning (using the same measure reported here [WSAS]) found a nonsignificant difference as compared to the control [62], a significant improvement for iCBT+treatment as usual but not in iCBT compared to the control (effect sizes not provided)

[52], and an effect size of 0.36 [29]. Two other studies found no significant difference [57] and an effect size of 0.65 [54] when measuring life/work functioning with the Sheehan Disability Scale. To our knowledge, no other iCBT studies have evaluated resilience. Lastly, this study found a nonsignificant reduction (58%) in the odds of having greater suicidal ideation in the Thrive intervention group than in the control group at 8 weeks. A small effect size (d=0.20) was found in the single prior study that measured the impact of an iCBT intervention on suicidality [58].

The design of this study supports the generalizability of its findings in community settings. First, the minimal eligibility restrictions indicate the potential applicability of the Thrive intervention in general populations outside the controlled research setting. Of the seven studies described in Karyotaki et al [21] and Lorenzo-Luaces et al [22], only one that delivered a fully automated, stand-alone iCBT intervention like the Thrive intervention had similarly broad eligibility criteria [29]. Second, omitting clinician-administered diagnostic evaluations, as used in several iCBT studies [57,59-61], allows for broader generalizability to community settings. The use of person-administered diagnostics greatly increases costs, limits broad intervention dissemination due to an insufficient mental health workforce, and decreases the potential public health impact of fully automated, stand-alone iCBT interventions.

Limitations

This study had several limitations. Like most iCBT studies [21], assessment completion rates were low, with 68% and 65% of randomized subjects completing assessments at 4 and 8 weeks, respectively. These rates are slightly below the 73% (range 55%-95%) unweighted mean completion rate [24-30,52,55,58-64] for short-term (6-16 weeks) follow-up assessments of all studies in Karyotaki et al [21]. Additionally, the study offered monetary incentives for 4- and 8-week survey completion, which are associated with greater response in electronic questionnaires [65] and retention rates [66], which limits conclusions about uptake of Thrive in nonresearch settings. Monetary incentives have been used in all six RCTs of iCBT interventions for depression in US populations [24,25,27,28,54,57] and three of the seven (43%) RCTs of fully automated, stand-alone iCBT interventions reported in Karyotaki et al and Lorenzo-Luaces et al [21,22,25,27,28]. Although using self-assessments, a common practice in iCBT studies, is a potential weakness of the study, the use of validated, widely used instruments largely addressed this issue. The PHQ-9 (depression symptoms), Generalized Anxiety Disorders Scale-7 (anxiety symptoms), and WSAS-5 (functioning) correlate well with clinician-administered instruments [37,67,68] and are sensitive to interventional effects [45,69,70]. Of note, self-assessments may underestimate the effect of iCBT relative to similar studies using clinician-administered assessments [71]. The study eligibility, which omitted the need for a depressive disorder diagnosis and included participants across the full range of depression symptom severity increases the study's generalizability. However, the make-up of the recruited study population, with high a proportion of female (85%) and white (93%) participants with at least some education after high school (69%) limits its generalizability. Greater participation of women is a common limitation among studies of iCBT for depression. For example, Karyotaki et al [21] reported that 66% of all participants were female, and among the studies of fully automated, stand-alone iCBT interventions, female participation ranged from 65% to 81% [21]. The high proportion of female participants across iCBT studies could be due, in part, to the greater prevalence of depression in women [2] or as a yet-unidentified barrier for men. The relatively high percentage (23.5%) of fraudulent participants identified is another limitation, which is becoming increasingly common in internet-based studies [72]. The number of fraudulent participants in this study is similar to that in several other studies using Web-based surveys: 28.7% [73], 20.5% [74], and 18.7% [75]. Although the study was powered appropriately to detect between-group differences in a priori defined primary and secondary outcomes, even when allowing for corrections for multiple testing (design strength of this study), it was underpowered to detect meaningful difference in the post-hoc assessment of suicidal ideation, as only 41% of subjects experienced suicidality at baseline. In addition, the 8-week evaluation period cannot inform on the long-term impact of the Thrive intervention when delivered in community settings. Longitudinal studies with longer follow-up periods are warranted to determine whether iCBT programs impact the risk of depression relapse [10]. Finally, as a community-based trial, these findings cannot be generalized to clinical settings.

Conclusions

Thrive, a fully automated, stand-alone iCBT intervention, demonstrated greater short-term improvements in depression and anxiety symptoms, work and social functioning, and resilience compared to control participants in a broad community population with mild to severe depression symptoms. A trend toward greater reduction in suicidal ideation was also observed. The magnitude of clinical benefit seen with the Thrive intervention appears to be greater than most other fully automated, stand-alone interventions and similar to those requiring clinician support. Thus, Thrive represents a potentially effective and cost-effective solution for treating depression symptoms in rural regions that often have few or no mental health providers. Evaluations of the impact of the Thrive intervention over longer periods of time and other population settings are warranted.

Acknowledgments

Research reported in this publication was supported by the National Institute of General Medical Sciences of the National Institutes of Health (P20GM103474, U54GM115371, and 5P20GM104417). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

At the time of the trial, JL and MB were affiliated with the Center for Mental Health Research and Recovery at Montana State University, Bozeman, Montana. JL is currently affiliated with the Department of Psychiatry at the University of Arizona, the Biomedical Research and Education Foundation of Southern Arizona, and the Southern Arizona VA Health Care System, Tucson, Arizona. MB is currently affiliated with the Department of Psychiatry at the University of Arizona and the Southern Arizona VA Health Care System, Tucson, Arizona.

Conflicts of Interest

JHG previously held a financial interest in Waypoint Health Innovations, which developed the Thrive intervention evaluated in this work. He no longer has a direct financial interest in Waypoint Health Innovations but does retain a small interest in Waypoint Health Innovations through Healthcare Technology Systems where he is CEO and a shareholder. Waypoint Health Innovations also pays him a royalty based on revenue from Thrive use. He is also a consultant to Waypoint on projects outside of the grant supporting this study. The terms of JHG's financial relationship with Waypoint Health Innovations have been reviewed by Montana State University, and his involvement with this research project has been approved in accordance with its conflict of interest policies.

Multimedia Appendix 1
Informed consent.
[DOCX File , 21 KB - jmir_v21i11e14754_app1.docx]

Multimedia Appendix 2
Description of study population with definition and removal of fraudulent participants.
[DOCX File , 30 KB - jmir_v21i11e14754_app2.docx]

Multimedia Appendix 3
CONSORT diagram.
[DOCX File , 64 KB - jmir_v21i11e14754_app3.docx]

Multimedia Appendix 4
Study safety protocol.
[DOCX File , 42 KB - jmir_v21i11e14754_app4.docx]

Multimedia Appendix 5
Covariates.
[DOCX File , 23 KB - jmir_v21i11e14754_app5.docx]

Multimedia Appendix 6
Power calculations.
[DOCX File , 34 KB - jmir_v21i11e14754_app6.docx]

Multimedia Appendix 7
Baseline PHQ-9 scores.
[DOCX File , 14 KB - jmir_v21i11e14754_app7.docx]

Multimedia Appendix 8
Adjusted least mean squares for secondary outcomes.
[PNG File , 539 KB - jmir_v21i11e14754_app8.png]

Multimedia Appendix 9
Effect of the Thrive intervention on suicidal ideation.
[DOCX File , 36 KB - jmir_v21i11e14754_app9.docx]

Multimedia Appendix 10
Participant flow chart.
[DOCX File , 78 KB - jmir_v21i11e14754_app10.docx]

Multimedia Appendix 11
CONSORT - EHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 352 KB - jmir_v21i11e14754_app11.pdf]

References

1. World Health Organization. 2017. Depression and other common mental disorders: global health estimates URL: https://apps.who.int/iris/bitstream/handle/10665/254610/WHO-MSD-MER-2017.2-eng.pdf [accessed 2019-11-08]
2. Hasin DS, Sarvet AL, Meyers JL, Saha TD, Ruan WJ, Stohl M, et al. Epidemiology of Adult DSM-5 Major Depressive Disorder and Its Specifiers in the United States. JAMA Psychiatry 2018 Apr 01;75(4):336-346 [FREE Full text] [doi: 10.1001/jamapsychiatry.2017.4602] [Medline: 29450462]
3. Gilman SE, Sucha E, Kingsbury M, Horton NJ, Murphy JM, Colman I. Depression and mortality in a longitudinal study: 1952-2011. CMAJ 2017 Oct 23;189(42):E1304-E1310 [FREE Full text] [doi: 10.1503/cmaj.170125] [Medline: 29061855]
4. Cuijpers P, Vogelzangs N, Twisk J, Kleiboer A, Li J, Penninx BW. Comprehensive meta-analysis of excess mortality in depression in the general community versus patients with specific illnesses. Am J Psychiatry 2014 Apr;171(4):453-462. [doi: 10.1176/appi.ajp.2013.13030325] [Medline: 24434956]
5. Berghöfer A, Roll S, Bauer M, Willich SN, Pfennig A. Screening for depression and high utilization of health care resources among patients in primary care. Community Ment Health J 2014 Oct;50(7):753-758. [doi: 10.1007/s10597-014-9700-4] [Medline: 24449430]
6. Chesney E, Goodwin GM, Fazel S. Risks of all-cause and suicide mortality in mental disorders: a meta-review. World Psychiatry 2014 Jun;13(2):153-160 [FREE Full text] [doi: 10.1002/wps.20128] [Medline: 24890068]
7. Biesheuvel-Leliefeld KEM, Kok GD, Bockting CLH, Cuijpers P, Hollon SD, van Marwijk HWJ, et al. Effectiveness of psychological interventions in preventing recurrence of depressive disorder: meta-analysis and meta-regression. J Affect Disord 2015 Mar 15;174:400-410. [doi: 10.1016/j.jad.2014.12.016] [Medline: 25553400]
8. Vittengl JR, Clark LA, Dunn TW, Jarrett RB. Reducing relapse and recurrence in unipolar depression: a comparative meta-analysis of cognitive-behavioral therapy's effects. J Consult Clin Psychol 2007 Jun;75(3):475-488 [FREE Full text] [doi: 10.1037/0022-006X.75.3.475] [Medline: 17563164]
9. Chand S, Chaitanya R, Chakkamparambil B, Prasad A, Vora A. CBT for depresssion: what the evidence says. Curr Psychiat 2018;17(9):14-23.
10. Webb CA, Rosso IM, Rauch SL. Internet-Based Cognitive-Behavioral Therapy for Depression: Current Progress and Future Directions. Harv Rev Psychiatry 2017;25(3):114-122 [FREE Full text] [doi: 10.1097/HRP.0000000000000139] [Medline: 28475503]
11. United States Census Bureau. 2016. New census data show differences between urban and rural populations URL: https://www.census.gov/newsroom/press-releases/2016/cb16-210.html [accessed 2019-06-01]
12. Kaiser Family Foundation. 2018 Sep 5. Mental Health Care Health Professional Shortage Areas (HPSAs) URL: https://tinyurl.com/ygtbh8w4 [accessed 2019-09-18]
13. US Department of Agriculture. State fact sheets: Montana URL: https://data.ers.usda.gov/reports.aspx?StateFIPS=30&StateName=Montana&ID=17854 [accessed 2019-09-18]
14. Centers for Disease Control and Prevention. 2017. BRFSS prevalence & trends data URL: https://www.cdc.gov/brfss/brfssprevalence [accessed 2019-09-18]
15. Centers for Disease Control and Prevention. 2018. Suicide Mortality by State URL: https://www.cdc.gov/nchs/pressroom/sosmap/suicide-mortality/suicide.htm [accessed 2019-09-18]
16. Andersson G, Cuijpers P. Internet-based and other computerized psychological treatments for adult depression: a meta-analysis. Cogn Behav Ther 2009 Dec;38(4):196-205. [doi: 10.1080/16506070903318960] [Medline: 20183695]
17. Andersson G, Cuijpers P, Carlbring P, Riper H, Hedman E. Guided Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: a systematic review and meta-analysis. World Psychiatry 2014 Oct;13(3):288-295 [FREE Full text] [doi: 10.1002/wps.20151] [Medline: 25273302]
18. Andrews G, Basu A, Cuijpers P, Craske MG, McEvoy P, English CL, et al. Computer therapy for the anxiety and depression disorders is effective, acceptable and practical health care: An updated meta-analysis. J Anxiety Disord 2018 Apr;55:70-78 [FREE Full text] [doi: 10.1016/j.janxdis.2018.01.001] [Medline: 29422409]
19. Vallury KD, Jones M, Oosterbroek C. Computerized Cognitive Behavior Therapy for Anxiety and Depression in Rural Areas: A Systematic Review. J Med Internet Res 2015;17(6):e139 [FREE Full text] [doi: 10.2196/jmir.4145] [Medline: 26048193]
20. Statista. 2017 Mar 11. U.S. home broadband penetration by state URL: https://www.statista.com/chart/10600/us-home-broadband-penetration-by-state/ [accessed 2019-09-18]
21. Karyotaki E, Riper H, Twisk J, Hoogendoorn A, Kleiboer A, Mira A, et al. Efficacy of Self-guided Internet-Based Cognitive Behavioral Therapy in the Treatment of Depressive Symptoms: A Meta-analysis of Individual Participant Data. JAMA Psychiatry 2017 Apr 01;74(4):351-359. [doi: 10.1001/jamapsychiatry.2017.0044] [Medline: 28241179]

22. Lorenzo-Luaces L, Johns E, Keefe JR. The Generalizability of Randomized Controlled Trials of Self-Guided Internet-Based Cognitive Behavioral Therapy for Depressive Symptoms: Systematic Review and Meta-Regression Analysis. J Med Internet Res 2018 Nov 09;20(11):e10113 [FREE Full text] [doi: 10.2196/10113] [Medline: 30413400]
23. Richards D, Richardson T. Computer-based psychological treatments for depression: a systematic review and meta-analysis. Clin Psychol Rev 2012 Jun;32(4):329-342. [doi: 10.1016/j.cpr.2012.02.004] [Medline: 22466510]
24. Gilbody S, Littlewood E, Hewitt C, Brierley G, Tharmanathan P, Araya R, et al. Computerised cognitive behaviour therapy (cCBT) as treatment for depression in primary care (REEACT trial): large scale pragmatic randomised controlled trial. BMJ 2015;351:h5627 [FREE Full text] [doi: 10.1136/bmj.h5627] [Medline: 26559241]
25. Clarke G, Reid E, Eubanks D, O'Connor E, DeBar LL, Kelleher C, et al. Overcoming depression on the Internet (ODIN): a randomized controlled trial of an Internet depression skills intervention program. J Med Internet Res 2002 Dec;4(3):E14 [FREE Full text] [doi: 10.2196/jmir.4.3.e14] [Medline: 12554545]
26. Kleiboer A, Donker T, Seekles W, van Straten A, Riper H, Cuijpers P. A randomized controlled trial on the role of support in Internet-based problem solving therapy for depression and anxiety. Behav Res Ther 2015 Sep;72:63-71. [doi: 10.1016/j.brat.2015.06.013] [Medline: 26188373]
27. Clarke G, Eubanks D, Reid E, Kelleher C, O'Connor E, DeBar LL, et al. Overcoming Depression on the Internet (ODIN) (2): a randomized trial of a self-help depression skills program with reminders. J Med Internet Res 2005;7(2):e16 [FREE Full text] [doi: 10.2196/jmir.7.2.e16] [Medline: 15998607]
28. Clarke G, Kelleher C, Hornbrook M, Debar L, Dickerson J, Gullion C. Randomized effectiveness trial of an Internet, pure self-help, cognitive behavioral intervention for depressive symptoms in young adults. Cogn Behav Ther 2009;38(4):222-234 [FREE Full text] [doi: 10.1080/16506070802675353] [Medline: 19440896]
29. Meyer B, Berger T, Caspar F, Beevers CG, Andersson G, Weiss M. Effectiveness of a novel integrative online treatment for depression (Deprexis): randomized controlled trial. J Med Internet Res 2009;11(2):e15 [FREE Full text] [doi: 10.2196/jmir.1151] [Medline: 19632969]
30. Mira A, Bretón-López J, García-Palacios A, Quero S, Baños RM, Botella C. An Internet-based program for depressive symptoms using human and automated support: a randomized controlled trial. Neuropsychiatr Dis Treat 2017;13:987-1006 [FREE Full text] [doi: 10.2147/NDT.S130994] [Medline: 28408833]
31. Montero-Marín J, Araya R, Pérez-Yus MC, Mayoral F, Gili M, Botella C, et al. An Internet-Based Intervention for Depression in Primary Care in Spain: A Randomized Controlled Trial. J Med Internet Res 2016;18(8):e231 [FREE Full text] [doi: 10.2196/jmir.5695] [Medline: 27565118]
32. Beck A, Rush A, Shaw B, Emery G. Cognitive Therapy Of Depression (the Guilford Clinical Psychology And Psychopathology Series). New York, NY: The Guilford Press; 1979.
33. Lewinsohn P, Munoz R, Youngren M, Zeiss A. Control Your Depression. Englewood Cliffs, NJ: Prentice-Hall Press; 1986.
34. Cuijpers P, van Straten A, Warmerdam L. Behavioral activation treatments of depression: a meta-analysis. Clin Psychol Rev 2007 Apr;27(3):318-326. [doi: 10.1016/j.cpr.2006.11.001] [Medline: 17184887]
35. Alberti R, Emmons M. Your Perfect Right: Assertiveness And Equality In Your Life And Relationships, 10th Edition. Oakland, CA: New Harbinger Publications; 2017.
36. Whiteside U, Richards J, Steinfeld B, Simon G, Caka S, Tachibana C, et al. Online cognitive behavioral therapy for depressed primary care patients: a pilot feasibility project. Perm J 2014 May;18(2):21-27 [FREE Full text] [doi: 10.7812/TPP/13-155] [Medline: 24867546]
37. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
38. United States Census Bureau. Quick facts: Montana URL: https://www.census.gov/quickfacts/MT [accessed 2019-09-18]
39. Montana State University. 2018. Thrive for Montana study URL: https://thriveformontana.com/ [accessed 2019-11-08]
40. Waypoint Health Innovations. Waypoint Health Innovations. 2018. Thrive URL: https://waypointhealth.com/thrive/ [accessed 2019-11-08]
41. Schure MB, Howard M, Bailey SJ, Bryan B, Greist J. Exploring Perceptions of a Computerized Cognitive Behavior Therapy Program in a U.S. Rural Western State. Rural Ment Health 2018 Jul;42(3-4):174-183. [doi: 10.1037/rmh0000102] [Medline: 30740195]
42. National Institute of Mental Health. 2019. Depression URL: https://www.nimh.nih.gov/health/topics/depression [accessed 2019-09-18]
43. National Alliance on Mental Illness - Montana. 2019. County Mental Illness Resource Guides URL: http://www.namimt.org/county-resource-guides.html [accessed 2019-09-18]
44. Löwe B, Decker O, Müller S, Brähler E, Schellberg D, Herzog W, et al. Validation and standardization of the Generalized Anxiety Disorder Screener (GAD-7) in the general population. Med Care 2008 Mar;46(3):266-274. [doi: 10.1097/MLR.0b013e318160d093] [Medline: 18388841]
45. Mundt JC, Marks IM, Shear MK, Greist JM. The Work and Social Adjustment Scale: a simple measure of impairment in functioning. Br J Psychiatry 2002 May;180:461-464. [doi: 10.1192/bjp.180.5.461] [Medline: 11983645]

46. Campbell-Sills L, Stein MB. Psychometric analysis and refinement of the Connor-davidson Resilience Scale (CD-RISC): Validation of a 10-item measure of resilience. J Trauma Stress 2007 Dec;20(6):1019-1028. [doi: 10.1002/jts.20271] [Medline: 18157881]
47. Wolfinger R, Chang M. Comparing the SAS GLM and MIXED procedures for repeated measures. Presented at: Proceedings of the Twentieth Annual SAS Users Group Conference; 1998; Cary, NC.
48. Little RJ. Modeling the Drop-Out Mechanism in Repeated-Measures Studies. Journal of the American Statistical Association 1995 Sep;90(431):1112-1121. [doi: 10.2307/2291350]
49. Rubin D. Inference and missing data. Biometrika 1976 Dec;63(3):581-592. [doi: 10.1093/biomet/63.3.581]
50. Documentation: United States Department of Agriculture. 2019. 2010 Rural-Urban Communiting Area (RUCA) Codes URL: https://www.ers.usda.gov/data-products/rural-urban-commuting-area-codes/documentation/ [accessed 2019-09-18]
51. Larson JE, Corrigan PW. Psychotherapy for self-stigma among rural clients. J Clin Psychol 2010 May;66(5):524-536. [doi: 10.1002/jclp.20679] [Medline: 20222119]
52. de Graaf LE, Gerhards SAH, Arntz A, Riper H, Metsemakers JFM, Evers SMAA, et al. Clinical effectiveness of online computerised cognitive-behavioural therapy without support for depression in primary care: randomised trial. Br J Psychiatry 2009 Jul;195(1):73-80. [doi: 10.1192/bjp.bp.108.054429] [Medline: 19567900]
53. Terides MD, Dear BF, Fogliati VJ, Gandy M, Karin E, Jones MP, et al. Increased skills usage statistically mediates symptom reduction in self-guided internet-delivered cognitive-behavioural therapy for depression and anxiety: a randomised controlled trial. Cogn Behav Ther 2018 Jan;47(1):43-61. [doi: 10.1080/16506073.2017.1347195] [Medline: 28724338]
54. Beevers CG, Pearson R, Hoffman JS, Foulser AA, Shumake J, Meyer B. Effectiveness of an internet intervention (Deprexis) for depression in a united states adult sample: A parallel-group pragmatic randomized controlled trial. J Consult Clin Psychol 2017 Apr;85(4):367-380. [doi: 10.1037/ccp0000171] [Medline: 28230390]
55. Farrer L, Christensen H, Griffiths KM, Mackinnon A. Internet-based CBT for depression with and without telephone tracking in a national helpline: randomised controlled trial. PLoS One 2011;6(11):e28099 [FREE Full text] [doi: 10.1371/journal.pone.0028099] [Medline: 22140514]
56. Berger T, Krieger T, Sude K, Meyer B, Maercker A. Evaluating an e-mental health program ("deprexis") as adjunctive treatment tool in psychotherapy for depression: Results of a pragmatic randomized controlled trial. J Affect Disord 2018 Feb;227:455-462. [doi: 10.1016/j.jad.2017.11.021] [Medline: 29154168]
57. Arean PA, Hallgren KA, Jordan JT, Gazzaley A, Atkins DC, Heagerty PJ, et al. The Use and Effectiveness of Mobile Apps for Depression: Results From a Fully Remote Clinical Trial. J Med Internet Res 2016 Dec 20;18(12):e330 [FREE Full text] [doi: 10.2196/jmir.6482] [Medline: 27998876]
58. Moritz S, Schilling L, Hauschildt M, Schröder J, Treszl A. A randomized controlled trial of internet-based therapy in depression. Behav Res Ther 2012 Aug;50(7-8):513-521. [doi: 10.1016/j.brat.2012.04.006] [Medline: 22677231]
59. Spek V, Nyklícek I, Smits N, Cuijpers P, Riper H, Keyzer J, et al. Internet-based cognitive behavioural therapy for subthreshold depression in people over 50 years old: a randomized controlled clinical trial. Psychol Med 2007 Dec;37(12):1797-1806. [doi: 10.1017/S0033291707000542] [Medline: 17466110]
60. Meyer B, Bierbrodt J, Schröder J, Berger T, Beevers CG, Weiss M, et al. Effects of an Internet intervention (Deprexis) on severe depression symptoms: Randomized controlled trial. Internet Interventions 2015 Mar;2(1):48-59. [doi: 10.1016/j.invent.2014.12.003]
61. Berger T, Hämmerli K, Gubser N, Andersson G, Caspar F. Internet-based treatment of depression: a randomized controlled trial comparing guided with unguided self-help. Cogn Behav Ther 2011;40(4):251-266. [doi: 10.1080/16506073.2011.616531] [Medline: 22060248]
62. Phillips R, Schneider J, Molosankwe I, Leese M, Foroushani PS, Grime P, et al. Randomized controlled trial of computerized cognitive behavioural therapy for depressive symptoms: effectiveness and costs of a workplace intervention. Psychol Med 2014 Mar;44(4):741-752 [FREE Full text] [doi: 10.1017/S0033291713001323] [Medline: 23795621]
63. Christensen H, Griffiths KM, Jorm AF. Delivering interventions for depression by using the internet: randomised controlled trial. BMJ 2004 Jan 31;328(7434):265 [FREE Full text] [doi: 10.1136/bmj.37945.566632.EE] [Medline: 14742346]
64. Klein JP, Berger T, Schröder J, Späth C, Meyer B, Caspar F, et al. Effects of a Psychological Internet Intervention in the Treatment of Mild to Moderate Depressive Symptoms: Results of the EVIDENT Study, a Randomized Controlled Trial. Psychother Psychosom 2016;85(4):218-228. [doi: 10.1159/000445355] [Medline: 27230863]
65. Brueton VC, Tierney JF, Stenning S, Meredith S, Harding S, Nazareth I, et al. Strategies to improve retention in randomised trials: a Cochrane systematic review and meta-analysis. BMJ Open 2014 Feb 04;4(2):e003821 [FREE Full text] [doi: 10.1136/bmjopen-2013-003821] [Medline: 24496696]
66. Liu Y, Pencheon E, Hunter RM, Moncrieff J, Freemantle N. Recruitment and retention strategies in mental health trials - A systematic review. PLoS One 2018;13(8):e0203127 [FREE Full text] [doi: 10.1371/journal.pone.0203127] [Medline: 30157250]
67. Löwe B, Kroenke K, Herzog W, Gräfe K. Measuring depression outcome with a brief self-report instrument: sensitivity to change of the Patient Health Questionnaire (PHQ-9). J Affect Disord 2004 Jul;81(1):61-66. [doi: 10.1016/S0165-0327(03)00198-8] [Medline: 15183601]

68. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
69. Zimmerman M, Walsh E, Friedman M, Boerescu DA, Attiullah N. Are self-report scales as effective as clinician rating scales in measuring treatment response in routine clinical practice? J Affect Disord 2018 Jan 01;225:449-452. [doi: 10.1016/j.jad.2017.08.024] [Medline: 28858659]
70. Kertz S, Bigda-Peyton J, Bjorgvinsson T. Validity of the Generalized Anxiety Disorder-7 scale in an acute psychiatric sample. Clin Psychol Psychother 2013;20(5):456-464. [doi: 10.1002/cpp.1802] [Medline: 22593009]
71. Cuijpers P, Li J, Hofmann SG, Andersson G. Self-reported versus clinician-rated symptoms of depression as outcome measures in psychotherapy research on depression: a meta-analysis. Clin Psychol Rev 2010 Aug;30(6):768-778. [doi: 10.1016/j.cpr.2010.06.001] [Medline: 20619943]
72. Teitcher JEF, Bockting WO, Bauermeister JA, Hoefer CJ, Miner MH, Klitzman RL. Detecting, preventing, and responding to "fraudsters" in internet research: ethics and tradeoffs. J Law Med Ethics 2015;43(1):116-133 [FREE Full text] [doi: 10.1111/jlme.12200] [Medline: 25846043]
73. Ballard AM, Cardwell T, Young AM. Fraud Detection Protocol for Web-Based Research Among Men Who Have Sex With Men: Development and Descriptive Evaluation. JMIR Public Health Surveill 2019 Feb 04;5(1):e12344 [FREE Full text] [doi: 10.2196/12344] [Medline: 30714944]
74. Bull S, Pratte K, Whitesell N, Rietmeijer C, McFarlane M. Effects of an Internet-based intervention for HIV prevention: the Youthnet trials. AIDS Behav 2009 Jun;13(3):474-487 [FREE Full text] [doi: 10.1007/s10461-008-9487-9] [Medline: 19037719]
75. Bauermeister J, Pingel E, Zimmerman M, Couper M, Carballo-Diéguez A, Strecher VJ. Data Quality in web-based HIV/AIDS research: Handling Invalid and Suspicious Data. Field methods 2012 Aug 01;24(3):272-291 [FREE Full text] [doi: 10.1177/1525822X12443097] [Medline: 23180978]

Abbreviations

CDRISC-10: Connor Davidson Resilience Scale-10
DSM: Diagnostic and Statistical Manual of Mental Disorders
GAD-7: Generalized Anxiety Disorder 7-item scale
iCBT: internet-based cognitive behavior therapy
LSM: least squares means
OR: odds ratio
PHQ-9: Patient Health Questionnaire-9
RCT: randomized controlled trial
WSAS-5: Work and Social Adjustment Scale-5

Edited by G Eysenbach; submitted 18.05.19; peer-reviewed by S Rauch, V Cosgrove, D Gratzer; comments to author 21.08.19; revised version received 23.09.19; accepted 22.10.19; published 18.11.19.

Please cite as:
Schure MB, Lindow JC, Greist JH, Nakonezny PA, Bailey SJ, Bryan WL, Byerly MJ
Use of a Fully Automated Internet-Based Cognitive Behavior Therapy Intervention in a Community Population of Adults With Depression Symptoms: Randomized Controlled Trial
J Med Internet Res 2019;21(11):e14754
URL: http://www.jmir.org/2019/11/e14754/
doi:10.2196/14754
PMID:31738173

Review

Young People's Online Help-Seeking and Mental Health Difficulties: Systematic Narrative Review

Claudette Pretorius[1], BSocSci, MA; Derek Chambers[2], MA; David Coyle[1], PhD

[1]School of Computer Science, University College Dublin, Dublin, Ireland

[2]Connecting for Life, Health Service Executive, Cork, Ireland

Corresponding Author:
Claudette Pretorius, BSocSci, MA
School of Computer Science
University College Dublin
Belfield
Dublin
Ireland
Phone: 353 017162818
Email: claudette.pretorius@ucdconnect.ie

Abstract

Background: Young people frequently make use of the internet as part of their day-to-day activities, and this has extended to their help-seeking behavior. Offline help-seeking is known to be impeded by a number of barriers including stigma and a preference for self-reliance. Online help-seeking may offer an additional domain where young people can seek help for mental health difficulties without being encumbered by these same barriers.

Objective: The objective of this systematic literature review was to examine young peoples' online help-seeking behaviors for mental health concerns. It aimed to summarize young peoples' experiences and identify benefits and limitations of online help-seeking for this age group. It also examined the theoretical perspectives that have been applied to understand online help-seeking.

Methods: A systematic review of peer-reviewed research papers from the following major electronic databases was conducted: PsycINFO, Cumulative Index of Nursing and Allied Health Literature, PubMed, Cochrane Library, Association for Computing Machinery Digital Library, and Institute of Electrical and Electronics Engineers Xplore. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines were followed. The search was conducted in August 2017. The narrative synthesis approach to reviews was used to analyze the existing evidence to answer the review questions.

Results: Overall, 28 studies were included. The most common method of data collection was through the use of surveys. Study quality was moderate to strong. Text-based query via an internet search engine was the most commonly identified help-seeking approach. Social media, government or charity websites, live chat, instant messaging, and online communities were also used. Key benefits included anonymity and privacy, immediacy, ease of access, inclusivity, the ability to connect with others and share experiences, and a greater sense of control over the help-seeking journey. Online help-seeking has the potential to meet the needs of those with a preference for self-reliance or act as a gateway to further help-seeking. Barriers to help-seeking included a lack of mental health literacy, concerns about privacy and confidentiality, and uncertainty about the trustworthiness of online resources. Until now, there has been limited development and use of theoretical models to guide research on online help-seeking.

Conclusions: Approaches to improving help-seeking by young people should consider the role of the internet and online resources as an adjunct to offline help-seeking. This review identifies opportunities and challenges in this space. It highlights the limited use of theoretical frameworks to help conceptualize online help-seeking. *Self-determination theory* and the *help-seeking model* provide promising starting points for the development of online help-seeking theories. This review discusses the use of these theories to conceptualize online help-seeking and identify key motivations and tensions that may arise when young people seek help online.

(J Med Internet Res 2019;21(11):e13873) doi:10.2196/13873

KEYWORDS

internet; help-seeking behavior; youth; mental health; online behavior; self-determination theory; systematic review

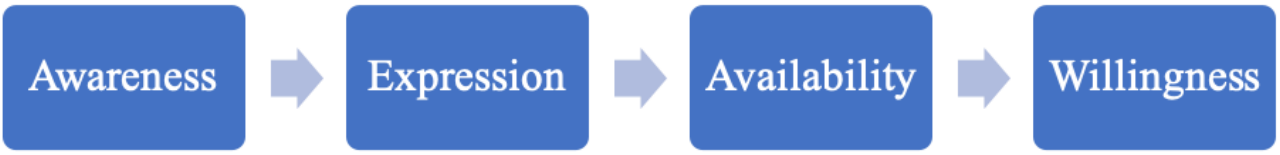
Introduction

Background

Mental health is an important health concern for young people around the world, with the World Health Organization estimating that 10% to 20% of young people experience mental health disorders [1]. It is estimated that 50% of all adult mental disorders start in adolescence [2]. Yet, most young people are reluctant to seek help from formal mental health services [3-5]. The help-seeking process is difficult; it is complicated by personal and contextual factors, such as access, stigma, and mental health literacy [3,6,7]. It then becomes critically important to find alternative methods in which to target and assist young people who are not receiving help.

Mental health help-seeking has been defined as "...an adaptive coping process that is the attempt to obtain external assistance to deal with a mental health concern" (p. 180) [8]. That external assistance can be from a wide array of sources. Rickwood et al [3] propose a conceptual model specifically taking into account the needs of young people (Figure 1). They describe help-seeking as a process that involves the following: (1) becoming aware of symptoms and making the appraisal that assistance might be required; (2) expressing the symptoms that they are experiencing and that they are in need of help or support; (3) the person should then be aware of sources of help that are available and accessible to them; and (4) the final step that depends on the willingness of the help-seeker to disclose their difficulties to the selected, available source [3].

Figure 1. Rickwood's Help-seeking model.



Two main types of help-seeking sources have been identified: formal and informal. Formal help-seeking can be understood as seeking assistance from any professional who has a recognized and legitimate role in providing support. Informal help-seeking is understood as pursuing assistance from informal social supports with whom the individual may or may not share a personal relationship [3]. Research shows that most young people have a preference for self-reliance when experiencing personal and emotional concerns and are more likely to make use of informal help sources than formal help sources, when and if they do reach out [3].

Most recently, computer-mediated technologies have begun to influence the help-seeking process. The internet offers another pathway to access care and help when young people are experiencing mental health concerns. Young people use the internet as their main source of information for all of their daily needs; accordingly, this logically extends to accessing information regarding their physical and mental health [9-11]. Various formal online services are readily available, as are informal resources such as discussion boards and social media [12]. Information gained from these sources could facilitate the help-seeking process to the next stage and could influence the way in which individuals form their help-seeking attitudes. The internet also offers unique benefits in the form of anonymity, access, and user control that can sometimes interfere with the offline help-seeking process [13]. The availability of high-quality mental health information and online resources could have a significant impact on the health outcomes of a young person [14].

Objectives

Although the potential benefits of online mental health resources have been acknowledged, there are also some concerns. For example, there is a worry that online help-seeking may delay access to formal help sources [10]. It is important to understand how these online resources, both formal and informal, are viewed by young people. Systematic reviews have been conducted in this area with a focus on how young people search for health-related information on the internet [15] and the effectiveness of online mental health services to improve help-seeking [14]; however, this review sought to understand the process and experiences of young people with regard to their online help-seeking experiences. The objectives of this narrative review [16] were to conduct a systematic analysis of the research on this topic and use the research to identify future opportunities for research and design that can improve the online help-seeking experiences of young people. The specific aims of this systematic review were as follows:

1. To examine the strategies employed by young people to search for help online for mental health difficulties.
2. To describe young people's experiences of online help-seeking for mental health difficulties.
3. To identify the benefits of young people's use of online mental health resources for help-seeking.
4. To identify the limitations of young people's use of online mental health resources for help-seeking.

Methods

Search Overview

This review was conducted adhering to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines and was registered on the PROSPERO database (PROSPERO registration number: CRD42017072487). On the basis of the aims of the study, inclusion and exclusion criteria were established to guide the subsequent search process.

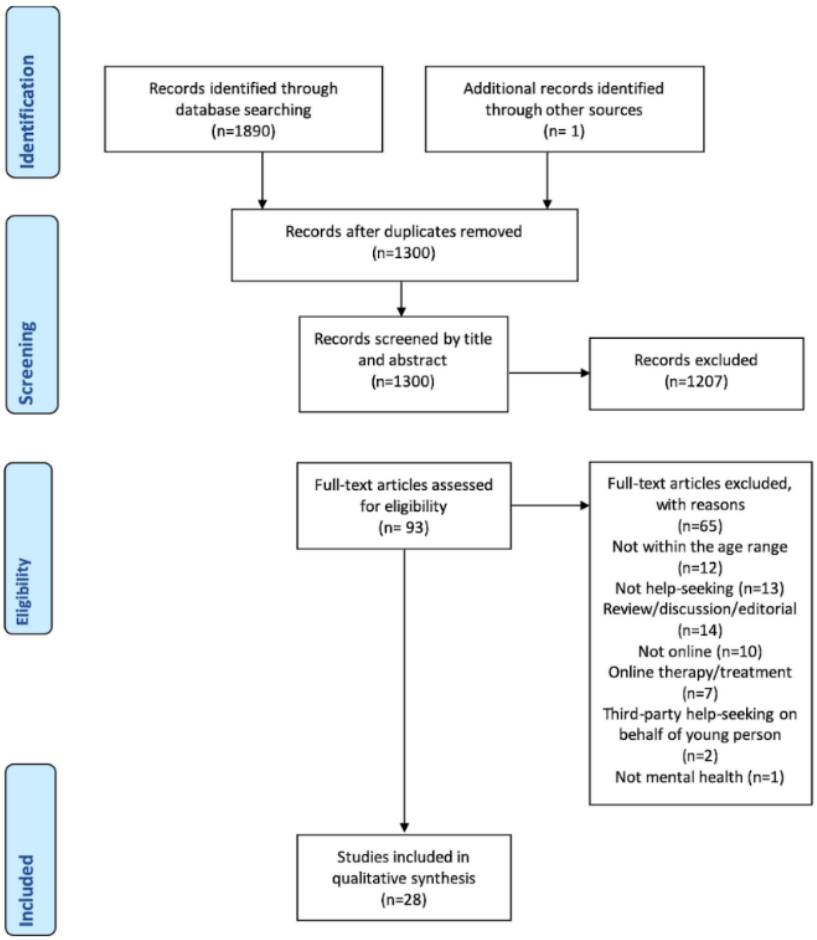
Search Strategy

The following 6 databases were searched from database inception (no limits were placed on the publication date as the evidence base in this area is very recent and to the author's (CP, DCUCD, and DCHSE) knowledge, a systematic review of this nature has not been completed before): PsycINFO, Cumulative Index of Nursing and Allied Health Literature, PubMed, Cochrane Library, Association for Computing Machinery Digital Library, and Institute of Electrical and Electronics Engineers Xplore in August 2017. In addition, the reference lists of all the included studies were scanned for relevant papers. The search terms aimed to represent the primary concepts of *online help-seeking*, *mental health*, and *young people*. Keywords were generated for each of these concepts by examining the terminology used in review papers in the help-seeking literature, and the authors sought the guidance of a trained librarian in the formation of the search string. The search strings are included in Multimedia Appendix 1. In keeping with the emerging youth mental health paradigm as described in the *International Declaration on Youth Mental Health* [17], the studies were restricted to young people aged 25 years and younger. Only English-language studies were included. All studies identified in the database search were exported to a reference managing software (EndNote X8 for Mac, Clarivate Analytics), and duplicate records were deleted.

The initial search identified 1890 published English-language abstracts (Figure 2). After removing the duplicates, 1300 papers remained. These papers were then reviewed by title and abstract to determine whether they met the inclusion criteria, resulting in 93 potentially relevant studies. At this stage, the full texts of these studies were obtained to confirm whether the inclusion and exclusion criteria listed below were met, resulting in 65 studies being excluded. The remaining 28 studies were included.

A random sample of 10% (130/1300, 10/93, 3/28) of the papers was re-examined at 3 stages (screening by title and abstract, screening by full text, and validity assessment) of the process by the other authors (DCUCD and DCHSE) of this paper. A few discrepancies were noted, and those were resolved by discussion and subsequent double checking to ensure consistency.

Figure 2. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRIMSA) flow diagram illustrating the screening process of papers.



Data Extraction

Each paper was read by CP and relevant details were extracted into a Microsoft Excel spreadsheet. There were a number of different protocol details that were examined and recorded for each paper. The coding scheme was used to help identify the components relevant to the study design and to address the research questions. The coding scheme included the year of publication, purpose of the study, country, number of participants, participants' characteristics (eg, medical conditions and age), theoretical framework, research design, sampling, data collection methods, instruments (including reliability and validity), data analysis, help-seeking model employed, strategies found to be used by young people to seek help online, young people's experiences of online help-seeking, facilitators, barriers, major findings, and study limitations.

Analysis and Synthesis

Quality assessment of included studies was conducted using the Critical Appraisal Skills Program (CASP) checklists [18] as there are quantitative and qualitative versions available to allow appraisal across different study designs. The CASP checklists enable the assessment of trustworthiness, relevance, and results of published papers and are divided into 3 sections to assess internal validity, results, and relevance to practice. Quality criteria for surveys included sections on research question and design, sampling framework and participant understanding, instrument metrics, response rate, coding and analysis, and result presentation [19-21].

These sections are assessed by questions that can be answered with *yes*, *no*, or *can't tell*. Each study received an overall rating of either *strong*, *moderate*, or *weak*, based on the number of questions scored as *yes*. Studies had to be scored as *yes* on majority of the questions to be rated *strong* overall (see Multimedia Appendix 2 tables for details, as criteria differed by study design).

The first author (CP) performed all of the quality assessments, with a 10% sample being given to the second author (DCUCD) to compare. Following the quality assessment stage, the inclusion of studies and extraction of key findings were finalized. Extracted data were entered into a table of study characteristics, including the quality assessment ratings for each study.

The decision was made to use narrative synthesis as it provides a broad overview of relevant information through a textual approach and is appropriate when it is expected that the studies will be heterogenous. Owing to the nature of the review questions, it was expected that the studies included would investigate help-seeking differently, make use of different research questions, and use different criteria to investigate help-seeking behaviors; thus, it would not be appropriate to make use of statistical synthesis techniques. The Guidance on the Conduct of Narrative Synthesis in Systematic Reviews [16] informed the data synthesis process to ensure that results and analysis were reported accurately. All the data extracted from the papers are presented narratively in text and summary tables.

Eligibility Criteria

The inclusion and exclusion criteria are shown in Textbox 1.

Textbox 1. Inclusion and exclusion criteria.

Inclusion criteria

- Young people not older than 25 years;
- Participants who present with psychological distress, self-selected general population samples, and those who have received a diagnosis from a health care practitioner;
- Studies designed to investigate and document the online help-seeking intentions and behaviors of young people;
- Studies that included an intervention or interventions that were designed to improve help-seeking attitudes or increase help-seeking intentions or help-seeking behaviors of young people;
- Web-based help-seeking interventions;
- Online mental health resources;
- Informal and formal help-seeking.

Exclusion criteria

- Studies focused on a third party seeking help for the young person;
- Not mental health related;
- Not a Web or mobile app–based;
- Unrelated technology, for example, computer games and their impact on mental health and social media and forums that are not specifically focused on mental health-related topics;
- Studies focused on online treatment methodologies and interventions, for example, computerized cognitive behavioral therapy or online counselling;
- A review paper.

Results

Study Characteristics

A total of 28 studies met the inclusion criteria (see Multimedia Appendix 3), of which the majority (n=16) were conducted in Australia. Others were conducted in the United Kingdom (n=4), Canada (n=4), the United States (n=2), Ireland (n=1), and the Netherlands (n=1). The studies were published between 2010 and 2017.

The research methodologies of the studies were varied. Only 4 of the 28 included studies mentioned which help-seeking model they employed to inform their study design [6,12,22,23]. These models were limited to Rickwood's *help-seeking model* and *the theory of planned behavior*. Survey research was the most widely employed research design; overall 11 studies [24-33] made use of an online survey and 4 studies [34-37] administered the survey in person to collect data from participants. Survey questions included standardized measures such as the General Help-Seeking Questionnaire and mental well-being scales such as the Kessler Psychological Distress Scale, but questions regarding technology and internet use varied and no standardized scales seemed to be available. Areas explored in these sections of the surveys included the internet and social media use, use of online mental health resources, perceived usefulness of these resources, and preference regarding these resources. Other study designs included 1 randomized controlled trial [38], 1 feasibility study [39], 2 comparative studies [40,41], and 4 qualitative studies [9,42-44]. These studies aimed to assess the intention to seek help from different help sources, previous help-seeking behaviors from different help sources, level of psychological distress, and preferred modes of delivery of online mental health services.

The number of participants in each study ranged from 23 to 3946. Their ages ranged from 12 to 25 years. The majority of studies had gender-mixed samples; however, 3 studies looked specifically at the online help-seeking behavior of males [9,22,25]. In all, 9 studies recruited participants who were students at a university or high school [9,22,24,27,30,31,35,37,42], whereas 6 studies [25,26,28,29,32,33] used online recruitment strategies. Most studies (n=22) were conducted with samples not selected on the basis of mental health status. However, 5 studies focused on participants who had reported experiencing self-harm and suicidal ideation [24,28,36,43,45], and 1 study specifically recruited participants who had been diagnosed with psychosis and nonpsychotic mood disorders [46].

Methodological Evaluation

Many of the studies achieved a strong or moderate rating on the checklists. However, the survey studies showed poor adherence or failed to report that they had indeed piloted the survey with a small number of young people before administering the surveys. Furthermore, randomized controlled trials showed small treatment effects, and there was poor evidence of rigorous data synthesis methods in the qualitative studies. See Multimedia Appendix 2.

Limitations of the Studies

All of the studies included some notable limitations which indicate specific gaps in the literature and findings that may not be generalizable to other populations. Many of the studies reported having participants, where the majority were female [27-31,33,35,40,45]. Although some studies focused specifically on males, it must be considered that much of the evidence in this area is from a female perspective. Many of the studies were based in Australia, where there is a great deal of investment in youth mental health services compared with other countries in the world. In addition, many of the studies recruited from a university population only [24,27,30,31,43], who are not representative of young people in general. Finally, many of the studies were cross-sectional or retrospective studies [12,24,29,36,47]; these types of studies include recall bias and do not accurately account for actual future behaviors.

How Young People Seek Help Online

In total, 6 studies found that young people made use of text-based queries using search engines to find mental health–related information rather than accessing a specific website [9,27,36,37,43,48] (see Multimedia Appendix 3). Information seeking about symptoms and forms of treatment are common goals when searching online [33]. The study by Birnbaum et al [47] showed that type of mental illness (mood disorder vs psychotic disorder) influenced what young people searched for. In this case, young people experiencing a mood disorder were more likely to search for how to stop symptoms, whereas those with a psychotic disorder preferred to understand why their symptoms had come about. When investigating the terms used by young people to search for mental health help, there was frequent use of *mental health*, *mental health problems*, *depression*, or *symptoms of...*, and *treatment of...* [30,37,43]. A common theme in the studies is the search for symptoms and treatment for the mental health concern the young person is currently facing [33]. The other types of content accessed that were identified by the studies include YouTube videos, factsheets, personal stories, and forums [33].

A total of 3 studies found that young people had used social media to locate mental health information [27,47,49]. The use of mental health or government websites varied from study to study. A study by Burns et al [48] found that less than 44.4% of the sample sourced information from mental health websites, similarly the study by Feng et al [27] found that only 26% of their sample had made use of information sites. Conversely, in a more recent study by Wetterlin et al [33], 82.9% of the sample indicated that they would be *somewhat likely* or *very likely* to use an information-based website. The study by Best et al [9] found that less of a quarter of their sample would make use of a government website.

Despite the lack of preference for formal mental health or government websites, a number of studies found that young people valued online services run by mental health professionals [9,28,47]. The study by Birnbaum et al [47] found that young people expressed an interest in obtaining help from mental health professionals through social media, whereas the study by Best et al [9] found that young people valued online services run by mental health professionals despite not wanting to use

government websites. Haner and Pepler [40] found that the more distress the young person was experiencing the more likely they were to access the *Live Chat* option with a website providing mental health support to young people. Similarly, Frost et al [28] found that young people who self-harm would prefer an online service that allowed them to directly link with a mental health professional via instant messaging when in crisis.

Online communities and discussion forums also serve as a platform young people use to seek help. In the analysis of an online community, Greidanus and Everall [44] found that help-seekers would come to the forum to post messages seeking help for personal distress, looking for input from other users of the online community. The use of discussion boards or online support groups was reported by 11% of the sample in the study by Feng and Campbell [27] and 48.6% of the sample in a study by Frost et al [28].

Motivating Factors for Young People to Seek Help Online

The studies identified by this review indicated that many young people were going online to look for a space where they could share their feelings without fear of judgement or labelling but at the same time, it was important that these spaces protected their privacy [30,33,43].

Many studies found that there is an association between high levels of psychological distress and engaging in help-seeking online [12,22,36,40,50]. Majority of this help-seeking tends to take place after 11 pm at night [22,48]. The study by Best et al [9] found that a preferred online service would be one run by professionals, available 24 hours a day. The need for services run and recommended by professionals is a recurrent theme throughout all of the studies [25,28,44]. A study by Wetterlin et al [33] found that 83.9% of participants reported that it was important to them to have human contact within an online mental health resource. This need for human contact also includes the need or desire to connect to peers online who can support the online help-seeking process; this is especially true of users of online support communities [30,33,37,41,43,44].

There seems to be a sentiment that there is more help available to young people online than offline, and that young people from minority groups and those with higher levels of psychological distress were more likely to disclose their current difficulties online rather than offline [32,35]. A study by Frost et al [28] indicated that this may be due to many young people finding online spaces to be less judgmental, and the support they received was nonstigmatizing. This is especially important because a number of studies found that those young people with increased levels of psychological distress were likely to access mental health content online but not seek help from offline sources [13,28,29,31,41].

Young People's Experiences of Online Help-Seeking

The findings regarding the perceived helpfulness of online resources were variable (see Table 1). Ellis et al [26] found that 81.9% of females in their sample reported that talking online had helped, and that they were either satisfied or very satisfied with the process. Similarly, 54.9% males had talked about their problems online, and 81.3% of men found that it had *helped* and were satisfied with the help they had received. The study by Feng and Campbell [27] found more mixed results with 59% of this sample indicating that the online resources they had made use of *didn't make things better or worse* and 40% reported that they had *helped a little*. Ruppel and McKinley [31] investigated social anxiety and levels of social support in relation to the perceived usefulness of online resources; they found that participants with higher levels of social anxiety and also those with high levels of social support found online support groups to be very useful. The analysis of an online community by Greidanus and Everall [44] found that young people had experienced the online communities (a community message board) as understanding and affirming; this sentiment was especially strong for users who felt misunderstood offline. Overall, the comments indicated that users had found their engagement on the site to be a positive experience.

Table 1. Findings identified in studies: what are young people's experiences of seeking help online?

Authors (year)	Findings related to young people's experiences online
Ellis et al (2013) [25]	• Most females said that talking online *helped* (81.9%), and that they were *satisfied* or *very satisfied* with the online help they received. More than half of all male respondents reported that they had talked about their problems online (54.9%). • Most said that talking online *helped* (81.3%), and that they were *satisfied* or *very satisfied* with the online help they received (82.9%).
Feng and Campbell (2011) [27]	• In total, 59% of participants reported that online resources that they had used *didn't make things better or worse*, 40% reported *they helped a little*, and only 1% of participants reported *they helped a lot*. • Although there is a preference for text-based search engines and information sites, the current sample does not seem to find them to be efficacious.
Frost and Casey (2016) [29]	• Over half of these online help-seekers perceived that they had more support available to them online than offline.
Frost et al (2016) [28]	• Young people identifying the need for a nonjudgmental (n=68) and safe (n=14) environment and interactions. Many young people used the term nonjudgmental, whereas others indicated that they needed support in a way that was not stigmatizing, did not stereotype them, blame them, or label them as an attention seeker. • Safety in online services for self-injury centered around the need for moderation, warnings about triggering content, and the risks of self-injury becoming competitive. • Young people with a previous experience of online help-seeking were more likely to endorse the importance of reduced isolation and a supportive online culture.
Greidanus and Everall (2010) [44]	• Most messages written by the trained volunteers took the form of an affirmation of some aspect of the help-seeker's character. • A strong sense of community was indicated in several of the threads when help-seekers stated they felt their experiences were understood and shared by other members. This sense appeared to be especially strong for those help-seekers who felt misunderstood by those in their *offline* lives. • Most of the community members authored a number of threads themselves and posted in threads of other members, occasionally making reference to the content of other threads. • Participant comments often indicated they found engagement on the site to be a positive experience and provided a place to express feelings, receive support, and obtain referrals.
Mars et al (2015) [36]	• Almost a quarter of the sample had come across a site that discussed self-harm or suicide.
Ruppel and McKinley (2015) [31]	• Participants with higher social support perceived websites and online support groups as more useful. • The perceived usefulness of online support groups was highest among participants who had high levels of social anxiety and high levels of social support.
Wetterlin et al (2014) [33]	• Most participants (87.7%) rated their privacy as a user as *very important*.

Benefits of Online Help-Seeking

Online help-seeking offers a number of benefits to young people experiencing personal and emotional difficulties. A total of 14 studies identified benefits of online help-seeking (see Table 2). These benefits could be grouped into 8 overarching categories which have been included in Table 3.

A total of 8 studies found that the anonymity provided by the internet was an important facilitator to online help-seeking [9,25,30,35,37,44]. Similarly ease of access and the immediacy of the internet plays an important role in its attractiveness to young people [9,12,28-30,35,51]. The nonstigmatizing nature of internet help-seeking makes it an attractive option for marginalized groups as seen in the study by Haner et al [40]. These groups include migrants and members of the LGBT+ community who may be fearful of disclosing personal concerns to their informal networks [40]. Similarly, a study by Best et al [22] found that online help-seeking was not affected by socioeconomic status or educational attainment.

Young people are finding a sense of community online and are able connect with others who have similar experiences to their own [24,43,46]. They feel they are able to communicate with this community without fear of judgment and, more importantly, they can control their level of disclosure [9,28]. In all, 2 studies indicated that young people who had previously gone online to seek help for self-injury or suicide-related issues were significantly less likely to have disclosed to someone offline [24,28].

Table 2. Benefits of seeking help online by study.

Authors (year)	Findings related to benefits of online help-seeking
Bell et al (2018) [24]	• Online help-seeking allows young people to communicate with others (social support but also reducing isolation). • Information is readily available. • Supportive sense of community and acceptance. • Comfort and relief in realizing that they are not alone.
Best et al (2016) [9]	• Anonymity • Ease of access • Immediacy • Absence of judgement • Can control level of disclosure
Best et al (2014) [22]	• Some males may not disclose problems to others, but they are receiving some form of support through help-seeking practices online. • Online help-seeking is not affected by Socio Economic Status or educational attainment. • Online sources may be providing young males with an additional outlet to seek social support.
Birnbaum et al (2017) [47]	• Opportunities for early intervention, as information found online can play an important role in the treatment-seeking decision-making process. • Young people are fearful to talk to close others about their symptoms but are comfortable to use the internet for further understanding. • Social media gives mental health clinicians the opportunity to engage and meaningfully interact with struggling youth at the earliest phases of illness potentially altering the trajectory to care. • Online information seeking plays an important role in the initiation of help-seeking by influencing individual's understanding of symptoms and their decision to seek professional help.
Bradford and Rickwood (2014) [35]	• Anonymity • Information that is easily accessible • Finding others who have similar experiences.
Burns et al (2016) [13]	• Reasons for preference of online resources included the anonymity of the internet, that information was easily accessible, and that there are often people in chat rooms who have been through the same thing. • Boys were shown to have a stronger preference for online resources compared with face-to-face help relative to girls.
Burns et al (2010) [48]	• Access online mental health resources in crisis outside of working hours (after 11 pm).
Collin et al (2011) [12]	• Online help-seeking helps young people to be more willing to ask a professional for help. • Upon having positive experience, help-seekers become advocates of help-seeking. • Gateway services promote timeous help-seeking.
Ellis et al (2013) [25]	• Preference instead for self-help and action-oriented strategies. • The internet addresses their desire for anonymity and self-help.
Frost and Casey (2016) [29]	• Online help-seekers indicated a greater intention to seek help for self-injurious behavior in the future. • A significant difference in help-seeking intentions from professionals emerged, with online help-seekers indicating significantly higher intentions to seek professional help compared with individuals who did not seek help online. • The internet may have an important role to play in mitigating help negation in young people who self-injure. • Young people who sought help online in relation to self-injury indicated a significantly greater intention to seek help for self-injurious behavior in the future, even after controlling for age, gender, and psychological distress.
Frost et al (2016) [28]	• Over half of the sample indicated a desire to use the internet as a first step but to later gain support offline. • The internet may provide a way of accessing support that is perceived as remaining private and within the control of the young person. • Perceived sense of community and belonging for young people who self-injure.

Authors (year)	Findings related to benefits of online help-seeking
Greidanus and Everall (2010) [44]	• It is clear these help-seekers, who reported not feeling comfortable seeking help from professional *offline* services, were able to use internet-based communication to create a community where they found support and offered support to their peers. • Children and adolescents who use alternative communication technologies find internet-based communications meaningful and personally relevant. • Help-seekers identified anonymity, accessibility, and access to peers who understand their experiences as important aspects of online help. • Help-seekers reported finding it easier to disclose some experiences online than offline.
Horgan and Sweeney (2010) [30]	• Anonymity, privacy, and confidentiality. • Accessibility, speed, and cost. • Believed that they would not be judged and believed it would be a good place to get initial information. • Easier to express themselves. • Ability to communicate with others in similar situations to find out how they are coping. • Young people indicated they are less likely to lie online. • Young people are reluctant to access mainstream mental health services because of fear of judgement and because of the stigma that still exists in relation to mental health problems.
Mar et al (2014) [43]	• Participants recounted using the internet to find others coping with similar problems, research their symptoms and prescribed medications, or understand their diagnosis. • Participants also emphasized that knowing there is a community of others helps them to recognize that they are not alone with their problems. • Participants sought a variety of e-mental health features, especially for engaging in active coping, such as journaling. • Online services may afford them a level of privacy. • E-mental health services may lessen the burden on providers or provide resources for patients waiting to access care. • E-mental health services may help to treat those with mild symptoms or those who do not wish to seek professional support. • Online services may also help direct those in need to the traditional health care system.

Table 3. Key benefit themes and number of studies in which each theme is addressed.

Serial no	Benefit	Number of studies	Studies
1	Anonymity and privacy	8	Best et al (2016) [9]; Bradford and Rickwood (2014) [35]; Burns et al (2016) [13]; Ellis et al (2013) [25]; Frost et al (2016) [28]; Greidanus and Everall (2010) [44]; Horgan and Sweeney (2010) [30]; Mar et al (2014) [43]
2	Ease of access and immediacy	7	Bell et al (2018) [24]; Best et al (2016) [9]; Bradford and Rickwood (2014) [35]; Burns et al (2016) [13]; Burns et al (2010) [48]; Greidanus and Everall (2010) [44]; Horgan and Sweeney (2010) [30]
3	Connecting with others with similar experiences	7	Bell et al (2018) [24]; Bradford and Rickwood (2014) [35]; Burns et al (2016) [13]; Frost et al (2016) [28]; Greidanus and Everall (2010) [44]; Horgan and Sweeney (2010); Mar et al (2014) [43]
4	Acts as a gateway to further help-seeking	5	Birnbaum et al (2017) [47]; Collin et al (2011) [12]; Frost and Casey (2016) [29]; Frost et al (2016) [28]; Mar et al (2014) [43]
5	Increased perceived control of help-seeking journey	3	Best et al (2016) [9]; Frost et al (2016) [28]; Mar et al (2014) [43]
6	Meets the needs of those with a preference for self-reliance	2	Ellis et al (2013) [25]; Mar et al (2014) [43]
7	Early access	2	Birnbaum et al (2017) [47]; Frost et al (2016) [28]
8	Inclusiveness of different Social Economic Status/cultures/genders	1	Best et al (2016) [22]

Online help-seeking seems to act as a gateway behavior to further help-seeking. It enables young people to access information about their mental health difficulties and, therefore, decide whether there is a need to seek professional help [9,12,28,46,49]. The internet provides alternative routes to access mental health professionals; for instance, Birnbaum et al [47] found that young people would be willing to access opportunities to connect with clinicians over social media. Collin et al [12] investigated the role of an online youth mental health website, ReachOut.com, in promoting young people's

help-seeking behavior. Users of the website (43.3% of those surveyed) indicated that using the website had helped them to acquire the skills and confidence to seek help if they needed it. Online resources could have a role to play in early intervention as the information found online could help early identification of concerning symptoms but also assist in reaching out to mental health professionals.

The internet also provides access to the information and tools that may assist those young people who have a preference for self-reliance or for informal sources of help [25,44]. In the study by Mar et al [43] participants indicated that they used the internet to search for active coping strategies such as journaling to assist them to cope with their current difficulties.

Limitations to Online Help-Seeking

Limitations to online help-seeking were discussed by 14 studies with 6 common categories of limitations found across all the studies (see Tables 4 and 5).

Young people are motivated to look for help online; however, their ability to access reliable, helpful information is influenced by their lack of mental health literacy and a lack of knowledge of which resources to search for [27,31,43,49]. In the context of help-seeking, mental health literacy can be understood as knowledge and understanding of mental health problems which aid their recognition and management [52]. Formal online services are limited and would need to be familiar to the help-seeker to be accessed [9]. A key concern raised by young people across 3 studies is that they were uncertain whether certain sources are reliable or not and lack an understanding of the indicators of quality [9,23,30]. It appears that young people attribute quality based on superficial characteristics such as rank on Google search results and design and layout of websites [9].

Treatment avoidance is a real risk associated with online help-seeking. Content exists online that can be stigmatizing, triggering, or that may reinforce harmful behaviors and thoughts [24,28]. Certain communities may also perpetuate the stigma surrounding mental health and psychiatric treatment options, which may contribute to a reluctance to seek help from offline, professional services [13,47]. The usual protective measures are not present in unmoderated communities, and it is concerning that risky content may not be removed [24]. This risk is exacerbated as young people may incorrectly attribute certain sources as *helpful* when in fact, they are dangerous [9,30,36,47]. Rickwood et al [41] expand on this by emphasizing that the self-reliance afforded by online help-seeking may have limited young people's access to the appropriate help source at the appropriate time because of limits of their own mental health literacy.

Finally, young people are concerned about the implications of making use of online help-seeking. These include fears over protection of privacy, that it may be too impersonal, and that the help found there would be unreliable and untrustworthy [30,39,43]. Mar et al [43] found that young people's concerns regarding their privacy centered around fears that family and friends would somehow find out about their mental health concern. A concern many of them also have about offline help-seeking.

Table 4. Themes identified in studies: limitations of seeking help online.

No.	Limitation	Number of studies	Studies
1	Uncertainty about trustworthiness of resources	5	Best et al (2016) [9]; Kauer et al (2017) [39]; Horgan and Sweeney (2010) [30];
2	Lack of mental health literacy	5	Bell et al (2018) [24]; Best et al (2016) [9]; Feng and Campbell (2011) [27]; Mar et al (2014) [43]; Ruppel and McKinley (2015) [31]
3	Reinforcing treatment avoidance	4	Birnbaum et al (2017) [47]; Mars et al (2015) [36]; Rickwood et al (2015) [41]
4	Concerns about privacy and confidentiality	3	Best et al (2016) [9]; Horgan and Sweeney (2010) [30]; Mar et al (2014) [43]
5	Triggering negative behavior	2	Bell et al (2018) [24]; Mars et al (2015) [36]
6	Difficulty in providing an emergency response	1	Mar et al (2014) [43]

Table 5. Limitations of seeking help online by study.

Author (year)	Findings related to limitations of online help-seeking
Bell et al (2018) [24]	• The risk of triggering or reinforcing suicidal thoughts or behaviors. • Unmoderated communities are risky as the fail-safe to remove risky content is not there.
Best et al (2016) [9]	• Lack of understanding of indicators of quality. • Lack of control of personal information once it is online. • Lack of confidentiality when disclosing within your own social network. • Lack of help-seekers' health literacy. • Formal online resources are limited and need to be known to be accessed.
Birnbaum et al (2017) [47]	• The online environment can be misleading and stigmatizing that reinforces pre-existing misconceptions about mental health and psychiatric treatment options, which may contribute to treatment avoidance.
Burns et al (2016) [13]	• There is still an overall orientation to not seek help, and barriers remain to all forms of help. This has concerning implications as it suggests that simply providing help through different means will not increase the likelihood that young people facing these barriers will actually use these new avenues of help.
Collin et al (2011) [12]	• Despite overall increased mental health literacy and intentions to seek help, ReachOut.com visitors remain reluctant to seek help from traditional and face-to-face sources.
Feng and Campbell (2011) [27]	• Young people are unaware of where to search for mental health concerns.
Frost et al (2016) [28]	• It is unclear whether online help-seeking was acting to replace offline help-seeking for these young people or whether the internet facilitates help-seeking in young people who otherwise would not disclose their self-injury to anyone. Similarly, it is unclear whether the failure of these young people to seek help offline may reflect a lack of linking to offline support in current forms of online support for self-injury. • Young people in the current sample went beyond discussion of the positive aspects of online communities and online culture, expressing concerns about triggering content, unmoderated discussions, and the *glorification* of self-injury.
Haner and Pepler (2016) [40]	• The possibility exists that the online counsellors can misinterpret neutral or positive typed communication with the presence of a vocal cue to suggest warmth of tone.
Horgan and Sweeney (2010) [30]	• A number of participants also reported that they believed it would be unreliable (15.1%), untrustworthy (5%), it lacks privacy (2.5%), is too impersonal (7.5%), and that insufficient support would be found (3.9%). • A number of participants were concerned with the reliability of the information, highlighting that young people may be experiencing difficulty in determining the quality of information online.
Kauer et al (2017) [39]	• Lack of trust in the accuracy of the information available on the internet was also a general concern for both the Link and comparison arms.
Mar et al (2014) [43]	• Two participants spoke of the importance of advertising the existence of online support offline, explaining that they felt it was not intuitive to look for help online. • Participants' concerns over privacy generally linked back to the stigma of having friends or family find out about their mental health concern. • Providers of e-mental health services for youth must appropriately address high suicide risk while maintaining a youth's privacy, which may need to be breached in emergency circumstances.
Mars et al (2015) [36]	• Young people have difficulties in classifying sites as either *helpful* or *harmful*, as some offer concurrent suicide-promoting and help-promoting content.
Rickwood et al (2015) [41]	• Greater self-reliance online, with a slightly stronger peer influence, may be cause for concern, as young people and their friends may not be the best guides to appropriate mental health care.
Ruppel and McKinley (2015) [31]	• Limited mental health literacy and limited knowledge about which resources are available.

Discussion

Principal Findings

This review aimed to extend understanding of how young people use online resources to seek help for their mental health concerns. A total of 28 studies were identified. Only 4 studies explicitly identified a theoretical framework of help-seeking that guided the study design. Moving forward, the development of such theoretical frameworks represents a key challenge. Results suggest that the internet serves 3 functions to help-seekers: (1) as a gateway to further information and

knowledge acquisition around their symptomology; (2) as a way to connect with others, professional or peer, around the topic of their mental health difficulties; and (3) as an alternative option to offline help-seeking for those who are most at risk. A text-based query via an internet search engine was the most commonly identified help-seeking approach. But social media, government or charity websites, live chat, instant messaging, and online communities and discussion forums are also used. The perceived benefits of online help-seeking include anonymity and privacy, ease of access, inclusivity, and the ability to connect with others and share experiences. Online help-seeking may also increase young peoples' sense of control over their help-seeking journey; meet the needs of those with a preference for self-reliance; or act as a gateway to further help-seeking. In contrast, significant limitations were also identified. A lack of mental health literacy can act as a barrier to effective help-seeking, as can concerns about privacy and confidentiality, and uncertainty about the trustworthiness of online resources. There is a concern that online help-seeking can reinforce treatment avoidance or trigger negative behavior.

Theoretical Frameworks in Online Help-Seeking

This review highlights the limited use of theoretical frameworks to help conceptualize online help-seeking and guide the development of improved resources. The full development of such a theory is beyond the scope of this paper. However, we consider 2 potential starting points for such a theory: Rickwood et al's *help-seeking model* [3] and *self-determination theory* (SDT) [53]. In each case, we use the existing theory as a lens through which to analyze the benefits and limitations of online help-seeking identified in this review and how they can either support or frustrate the help-seeking process.

The Help-Seeking Model

Existing theories of help-seeking provide a valuable starting point for the development of online help-seeking theories. The *help-seeking model* of Rickwood et al [3] provides a stage-based model to understand traditional help-seeking behaviors. It was applied by 2 studies in this review. Best et al [9] have proposed a pathways-based extension of this theory through their *pathways to online help-seeking model*, which seeks to predict people's help-seeking decisions on the basis of their mental health literacy and perception of stigma.

Table 6 demonstrates another approach to applying the *help-seeking model*. It outlines 1 way in which the benefits and limitations identified in this review can be mapped to stages of the *help-seeking model*. Through such a mapping, we can begin to identify and think about important issues that impact different stages of an online help-seeking process. For example, early access and the potential of online services to act as a gateway to further help-seeking may offer significant benefit at the awareness stage of a help-seeking process. This is offset by the lack of mental health literacy many young people will have at this stage of the process. Similarly, a lack of literacy is also likely to impact the expression stage. However, while young people may struggle to recognize or express their symptoms using formal clinical language, they might benefit from reading the stories of other young people, whose experiences they might relate to, potentially helping them to understand their own symptoms in a more accessible manner, and thus enabling expression.

As shown in Table 6, this approach can also be applied at the availability and willingness stages. In each case the approach provides a structured way to think about benefits we might maximize, while also highlighting limitations that need to be addressed. Such an analysis can guide the design of more effective online help-seeking services. It is also important to note the we do not see the mapping presented here as exclusive. We recognize that other mappings are possible. Our intention is to demonstrate how consideration of distinct benefits and limitations at each stage of a help-seeking model can shed light on key challenges and opportunities in the design of online help-seeking services.

Table 6. A mapping of the benefits and limitations to online help-seeking based on the stages of *help-seeking model* by Rickwood et al [3].

Stage	Awareness	Expression	Availability	Willingness
Process	Becoming aware of symptoms, appraising the assistance required	Expressing the symptoms experienced and that they are in need of help or support	Identify sources of help that are available and accessible	Willingness of the help-seeker to disclose difficulties to the selected, available source
Benefit/support	• Early access • Acts as a gateway to further help-seeking	• Connecting with others with similar experiences	• Ease of access and immediacy • Inclusiveness • Meets the needs of those with a preference for self-reliance	• Anonymity and privacy • Control of help-seeking journey • Connecting with others with similar experiences
Limitation/frustration	• Lack of mental health literacy	• Lack of mental health literacy	• Lack of immediate, crisis support	• Concerns about privacy and confidentiality • Treatment avoidance • Triggering negative behavior

Self-Determination Theory

SDT is a theory of motivation that has been applied across many settings in education and health care to understand and predict psychological well-being [53,54]. In recent years, it has also been applied in the design of digital technologies that can support mental health and well-being [55,56]. Until now, it has not been applied to online help-seeking.

SDT consists of a number of mini-theories; one of which is *cognitive evaluation theory* [53,54]. This theory proposes that there are 3 primary psychological needs for well-being and motivation: autonomy (to experience choice in line with one's own interests and values); competence (effectively interact with one's environment and express one's abilities); and relatedness (sense of belonging) [57]. SDT postulates that these basic 3 needs are essential for understanding how and why humans pursue certain goals. It asserts that the natural human trajectory is toward vitality, integration, and health [58] and argues that environments can support or frustrate these needs and thus influence well-being and motivation.

Here, we consider how SDT can be applied to conceptualize motivation in online help-seeking. In Table 7, the benefits and limitations of online help-seeking identified in this review are clustered in terms of their impact on autonomy, competence, and relatedness as consistent with SDT. In online help-seeking, the goal might vary from person to person; however, those specific goals could be understood within the broad themes of achieving growth and well-being. One could, therefore, argue that online searches and resources that are designed to support these basic psychological needs will be better able to support young people when they engage in help-seeking.

The evidence from this review suggests that there were mixed responses with regard to young people's satisfaction with their experiences when looking for help online. This mixed response could be attributed to online resources not fully meeting the needs for autonomy, competence, and relatedness, or these needs only being partially met. An example of this could be the use of a text-based search engine, which facilitates the need for autonomy, but leads to an abundance and variety of search results that could overwhelm the young person, frustrating their need for competence. The mapping also shows how a lack of mental health literacy can be thought of as an issue of competence. Addressing this competence may increase overall motivation for help-seeking. Similarly, the decision to avoid treatment may be a negatively focused expression of autonomy. In such a conceptualization, we might predict that systems which support alternative forms of autonomy, for example, by making control of the journey more explicit, will reduce the likelihood of treatment avoidance.

The importance of connecting with others online, whether professionals or peers, is emphasized by participants in many of the reviewed studies, indicating the importance for a human element in both formal and informal online help sources. Previous research has also shown that knowing someone who has sought help for a mental health difficulty has a positive effect on one's attitude toward help-seeking [3,53,54]. Viewed through the lens of SDT, engaging with people online or with content that shares the stories of other's help-seeking journeys helps to provide relatedness and improve mental health literacy. As such the internet may play an important bridging role between different stages of a help-seeking process, first facilitating informal contact, but also increasing motivation toward formal help-seeking.

Table 7. A clustering of the benefits and limitations to online help-seeking on the basis of the primary psychological needs identified in self-determination theory.

Benefits and Limitations	Autonomy	Competence	Relatedness
Benefit/support	• Anonymity and privacy • Ease of access and immediacy • Control of help-seeking journey • Meets the needs of those with a preference for self-reliance	• Acts as a gateway to further help-seeking • Early access	• Connecting with others with similar experiences • Inclusiveness
Limitation/frustration	• Concerns about privacy and confidentiality • Treatment avoidance	• Lack of mental health literacy	• Lack of immediate, crisis support • Triggering negative behavior

Tensions and Opportunities in Online Help-Seeking

Evidence suggests that the internet has the potential to serve as an inclusive gateway that assists all young people in accessing help, especially those from minority groups and groups who experience a great deal of stigma. It provides immediacy of access and allows people to connect with others, while also preserving the option to remain anonymous and control how much information they reveal, thereby supporting relatedness and also meeting the need for autonomy. However, many young people may feel forced to limit what they reveal or how they search due to concerns around privacy and confidentiality, now limiting their autonomy. This highlights a tension between the potential benefit of human contact versus the need for confidentiality. Relatedly, there is a tension between the preference of some young people for self-reliance versus the benefits of disclosure to formal sources of help.

Rickwood et al's model emphasizes the importance of the social transactions implicit in traditional offline help-seeking [3]. However, online help-seeking changes the nature of and need for social transactions with others. This review has found that while some online help-seekers prefer online mental health resources that offer the opportunity to connect with others, others prefer to navigate the process on their own and rely on self-help strategies, removing the need for a social transaction. The different types of content on the internet mean that the nature of social transactions has also changed: a help-seeker can now read content regarding another young person's personal experiences of mental health difficulties without ever directly engaging with the original writer of the content. The writer of the content can still have a profound effect on the help-seeker, not only improving their mental health literacy but also providing an example of someone who has also sought help for

a mental health difficulty. In this way, help-seekers become the consumers as well as the creators of help-seeking content online.

Online help-seeking can facilitate young people's autonomy by allowing them to control their help-seeking journey. However, as noted, young people have differing preferences for which online resources they access and which resources they find useful. These preferences are not just different from person to person. They may also differ for any given person, depending on their immediate circumstances. A key challenge for future research lies in providing tailored and appropriate online resources for different preferences and groups that meet all 3 psychological needs. Addressing all 3 needs in an online resource currently appears to be lacking. For example, young people are accessing resources where they can read and share personal stories, which meets their need for relatedness but it is unlikely that they would meet their needs for competence (is this information trustworthy?) and autonomy (is my privacy ensured?) through these resources.

Our analysis suggests that young peoples' online help-seeking may trigger key tensions between supporting and frustrating the 3 basic psychological needs outlined in SDT; simultaneously, the internet and online resources have the opportunity and capability to address these needs through careful and considered design. Managing these tensions will be important if we are to realize the full potential of systems that support online help-seeking.

Implications for Practice

As the internet becomes increasingly a part of everyday life and is seen as an accessible tool for information, it is important to have an understanding of how young people use the internet to meet their mental health needs. A plethora of online resources exist, both good and bad, and we only have a limited understanding of the patterns and characteristics of young people's mental health–related internet use. Online help-seeking provides an added space for young people to access help sources; however, it is an addition to offline help-seeking and not a replacement. There remains a great need to educate young people to facilitate competent and appropriate help-seeking behavior, both online and offline. Online sources need to be designed with young people's needs in mind, specifically making services available after-hours and providing access to trained professionals and peers. To increase reach, offline service providers need to consider online/digital strategies to offer a continuum of services to address the mental health needs of young people. Similarly, those developing online resources for young people need to do so in collaboration with professionals and young people.

Limitations of This Review

This review has several limitations. Although a number of databases have been included, the choice of keywords may have resulted in missing relevant research. Owing to the exploratory nature of this review, the decision was made to include a wide range of study designs, and the review will ultimately be limited by the design of the studies included. Although strategies to limit bias were included through consultation with the second and third reviewer, the possibility of subjectivity in analyzing the findings is acknowledged. Additionally, the measures used in the studies were varied and samples were heterogenous, making it a challenge to compare outcomes across studies. It is also evident from the studies included in this review that further investigation is needed into the online help-seeking behaviors of young people from populations other than those included in these studies. These data are representative of a mostly female, university student sample, often living in Australia. These findings may not translate well onto other populations such as those young people living in Europe, Asia, or Africa, young men and those young people who have not accessed tertiary education.

Conclusions

A key concern for researchers in this area should be the development of a model or framework in which to explain the motivations and benefits of online help-seeking and how it fits in with the overall help-seeking process for young people. The conceptualization of such a model would contribute to the cross-validation of findings and provide the ability to determine patterns. This review has considered *the help-seeking model* and SDT as valuable starting points for such a theory. It would allow research questions to be framed within the SDT constructs and theories, allowing for comparison and validation.

Acknowledgments

CP is supported by a grant from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No. 722561.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search Strings.
[PDF File (Adobe PDF File), 71 KB - jmir_v21i11e13873_app1.pdf]

Multimedia Appendix 2
Quality Checklists.

[PDF File (Adobe PDF File), 129 KB - jmir_v21i11e13873_app2.pdf]

Multimedia Appendix 3
Tables 1 and 2.
[PDF File (Adobe PDF File), 153 KB - jmir_v21i11e13873_app3.pdf]

References

1. WHO. Investing in Mental Health. Geneva: World Health Organization; 2003.
2. Kessler RC, Amminger GP, Aguilar-Gaxiola S, Alonso J, Lee S, Ustün TB. Age of onset of mental disorders: a review of recent literature. Curr Opin Psychiatry 2007 Jul;20(4):359-364 [FREE Full text] [doi: 10.1097/YCO.0b013e32816ebc8c] [Medline: 17551351]
3. Rickwood D, Deane F, Wilson C, Ciarrochi JV. Young people's help-seeking for mental health problems. Aust J Adv Mental Heal 2005;4(3):1-34. [doi: 10.5172/jamh.4.3.218]
4. Rowe SL, French RS, Henderson C, Ougrin D, Slade M, Moran P. Help-seeking behaviour and adolescent self-harm: a systematic review. Aust N Z J Psychiatry 2014 Dec;48(12):1083-1095. [doi: 10.1177/0004867414555718] [Medline: 25335872]
5. Gulliver A, Griffiths KM, Christensen H. Perceived barriers and facilitators to mental health help-seeking in young people: a systematic review. BMC Psychiatry 2010 Dec 30;10:113 [FREE Full text] [doi: 10.1186/1471-244X-10-113] [Medline: 21192795]
6. Cunningham CE, Walker JR, Eastwood JD, Westra H, Rimas H, Chen Y, The Mobilizing Minds Research Group. Modeling mental health information preferences during the early adult years: a discrete choice conjoint experiment. J Health Commun 2014 Apr;19(4):413-440 [FREE Full text] [doi: 10.1080/10810730.2013.811324] [Medline: 24266450]
7. Daine K, Hawton K, Singaravelu V, Stewart A, Simkin S, Montgomery P. The power of the web: a systematic review of studies of the influence of the internet on self-harm and suicide in young people. PLoS One 2013;8(10):e77555 [FREE Full text] [doi: 10.1371/journal.pone.0077555] [Medline: 24204868]
8. Rickwood D, Thomas K, Bradford S. The Sax Institute. 2012. Help-seeking Measures in Mental Health: A Rapid Review URL: https://www.saxinstitute.org.au/wp-content/uploads/02_Help-seeking-measures-in-mental-health.pdf [accessed 2019-10-02]
9. Best P, Gil-Rodriguez E, Manktelow R, Taylor B. Seeking help from everyone and no-one: conceptualizing the online help-seeking process among adolescent males. Qual Health Res 2016 Jul;26(8):1067-1077. [doi: 10.1177/1049732316648128] [Medline: 27179020]
10. Gowen LK. Online mental health information seeking in young adults with mental health challenges. J Technol Hum Serv 2013;31(2):97-111. [doi: 10.1080/15228835.2013.765533]
11. Subramaniam M, Taylor NG, Jean B, Follman R, Kodama C, Casciotti D. As simple as that?: tween credibility assessment in a complex online world. J Doc 2015;71(3):550-571. [doi: 10.1108/JD-03-2014-0049]
12. Collin P, Metcalf A, Stephens-Reicher J, Blanchard M, Herrman H, Rahilly K, et al. ReachOut.com: The role of an online service for promoting help-seeking in young people. Adv Mental Heal 2011;10(1):39-51. [doi: 10.5172/jamh.2011.10.1.39]
13. Burns JM, Birrell E, Bismark M, Pirkis J, Davenport TA, Hickie IB, et al. The role of technology in Australian youth mental health reform . Aust Health Rev 2016 Nov;40(5):584-590. [doi: 10.1071/AH15115] [Medline: 26934382]
14. Kauer SD, Mangan C, Sanci L. Do online mental health services improve help-seeking for young people? A systematic review. J Med Internet Res 2014 Mar 4;16(3):e66 [FREE Full text] [doi: 10.2196/jmir.3103] [Medline: 24594922]
15. Park E, Kwon M. Health-related internet use by children and adolescents: systematic review. J Med Internet Res 2018 Apr 3;20(4):e120 [FREE Full text] [doi: 10.2196/jmir.7731] [Medline: 29615385]
16. Dreyer NA, Velentgas P, Westrich K, Dubois R. The GRACE checklist for rating the quality of observational studies of comparative effectiveness: a tale of hope and caution. J Manag Care Spec Pharm 2014 Mar;20(3):301-308 [FREE Full text] [doi: 10.18553/jmcp.2014.20.3.301] [Medline: 24564810]
17. Indivisible - Squarespace. The International Declaration on Youth Mental Health URL: https://static1.squarespace.com/static/58d451526a49630cabca750d/t/5b5dc8fd758d46f72ceacd5d/1532872966444/8909_Int+Declaration+YMH_print.pdf [accessed 2019-10-02]
18. Critical Appraisal Skills Programme. URL: https://casp-uk.net/ [accessed 2019-02-28] [WebCite Cache ID 76WSPlq1M]
19. Rolls K, Hansen M, Jackson D, Elliott D. How health care professionals use social media to create virtual communities: an integrative review. J Med Internet Res 2016 Jun 16;18(6):e166 [FREE Full text] [doi: 10.2196/jmir.5312] [Medline: 27328967]
20. Ponto J. Understanding and evaluating survey research. J Adv Pract Oncol 2015;6(2):168-171 [FREE Full text] [Medline: 26649250]
21. Wang L, Fan X. Six Criteria for Survey Sample Design Evaluation. In: Paper presented at the Annual Meeting of the American Educational Research Association. 1998 Presented at: AERA'98; April 13-17, 1998; San Diego, CA.

22. Best P, Manktelow R, Taylor BJ. Social work and social media: online help-seeking and the mental well-being of adolescent males. Br J Soc Work 2014;46(1):257-276. [doi: 10.1093/bjsw/bcu130]
23. Kauer SD, Buhagiar K, Sanci L. Facilitating mental health help-seeking in young adults: the underlying theory and development of an online navigation tool. Adv Mental Heal 2016;15(1):71-87. [doi: 10.1080/18387357.2016.1237856]
24. Bell J, Mok K, Gardiner E, Pirkis J. Suicide-related internet use among suicidal young people in the UK: characteristics of users, effects of use, and barriers to offline help-seeking. Arch Suicide Res 2018;22(2):263-277. [doi: 10.1080/13811118.2017.1334609] [Medline: 28590842]
25. Ellis LA, Collin P, Hurley PJ, Davenport TA, Burns JM, Hickie IB. Young men's attitudes and behaviour in relation to mental health and technology: implications for the development of online mental health services. BMC Psychiatry 2013 Apr 20;13:119 [FREE Full text] [doi: 10.1186/1471-244X-13-119] [Medline: 23601273]
26. Ellis LA, Collin P, Davenport TA, Hurley PJ, Burns JM, Hickie IB. Young men, mental health, and technology: implications for service design and delivery in the digital age. J Med Internet Res 2012 Nov 22;14(6):e160 [FREE Full text] [doi: 10.2196/jmir.2291] [Medline: 23171827]
27. Feng XL, Campbell A. Understanding e-mental health resources: personality, awareness, utilization, and effectiveness of e-mental health resources amongst youth. J Technol Hum Serv 2011;29(2):101-119. [doi: 10.1080/15228835.2011.595276]
28. Frost M, Casey L, Rando N. Self-injury, help-seeking, and the internet: informing online service provision for young people. Crisis 2016;37(1):68-76. [doi: 10.1027/0227-5910/a000346] [Medline: 26572908]
29. Frost M, Casey L. Who seeks help online for self-injury? Arch Suicide Res 2016;20(1):69-79. [doi: 10.1080/13811118.2015.1004470] [Medline: 25706352]
30. Horgan A, Sweeney J. Young students' use of the internet for mental health information and support. J Psychiatr Ment Health Nurs 2010 Mar;17(2):117-123. [doi: 10.1111/j.1365-2850.2009.01497.x] [Medline: 20465756]
31. Ruppel EK, McKinley CJ. Social support and social anxiety in use and perceptions of online mental health resources: exploring social compensation and enhancement. Cyberpsychol Behav Soc Netw 2015 Aug;18(8):462-467. [doi: 10.1089/cyber.2014.0652] [Medline: 26252931]
32. Spears BA, Taddeo CM, Daly AL, Stretton A, Karklins LT. Cyberbullying, help-seeking and mental health in young Australians: implications for public health. Int J Public Health 2015 Feb;60(2):219-226. [doi: 10.1007/s00038-014-0642-y] [Medline: 25572385]
33. Wetterlin FM, Mar MY, Neilson EK, Werker GR, Krausz M. eMental health experiences and expectations: a survey of youths' web-based resource preferences in Canada. J Med Internet Res 2014 Dec 17;16(12):e293 [FREE Full text] [doi: 10.2196/jmir.3526] [Medline: 25519847]
34. Birnbaum ML, Garrett C, Baumel A, Scovel M, Rizvi AF, Muscat W, et al. Using digital media advertising in early psychosis intervention. Psychiatr Serv 2017 Nov 1;68(11):1144-1149. [doi: 10.1176/appi.ps.201600571] [Medline: 28712355]
35. Bradford S, Rickwood D. Adolescent's preferred modes of delivery for mental health services. Child Adolesc Ment Health 2014;19(1):39-45. [doi: 10.1111/camh.12002]
36. Mars B, Heron J, Biddle L, Donovan JL, Holley R, Piper M, et al. Exposure to, and searching for, information about suicide and self-harm on the Internet: prevalence and predictors in a population based cohort of young adults. J Affect Disord 2015 Oct 1;185:239-245 [FREE Full text] [doi: 10.1016/j.jad.2015.06.001] [Medline: 26150198]
37. O'Dea B, Campbell A. Healthy connections: online social networks and their potential for peer support. Stud Health Technol Inform 2011;168:133-140. [doi: 10.3233/978-1-60750-791-8-133] [Medline: 21893921]
38. Sanci LA, Kauer S, Buhagiar K. A randomised controlled trial of an online mental health help-seeking tool for young people: the link project. J Adolesc Health 2017;60(2):S4. [doi: 10.1016/j.jadohealth.2016.10.030]
39. Kauer SD, Buhagiar K, Blake V, Cotton S, Sanci L. Facilitating mental health help-seeking by young adults with a dedicated online program: a feasibility study of Link. BMJ Open 2017 Jul 9;7(7):e015303 [FREE Full text] [doi: 10.1136/bmjopen-2016-015303] [Medline: 28694345]
40. Haner D, Pepler D. 'Live chat' clients at kids help phone: individual characteristics and problem topics. J Can Acad Child Adolesc Psychiatry 2016;25(3):138-144 [FREE Full text] [Medline: 27924143]
41. Rickwood DJ, Mazzer KR, Telford NR. Social influences on seeking help from mental health services, in-person and online, during adolescence and young adulthood. BMC Psychiatry 2015 Mar 7;15:40 [FREE Full text] [doi: 10.1186/s12888-015-0429-6] [Medline: 25886609]
42. Havas J, de Nooijer J, Crutzen R, Feron F. Adolescents' views about an internet platform for adolescents with mental health problems. Health Educ 2011;111(3):164-176. [doi: 10.1108/09654281111123466]
43. Mar MY, Neilson EK, Torchalla I, Werker GR, Laing A, Krausz M. Exploring e-Mental health preferences of generation y. J Technol Hum Serv 2014;32(4):312-327. [doi: 10.1080/15228835.2014.943457]
44. Greidanus E, Everall RD. Helper therapy in an online suicide prevention community. Br J Guid Counc 2010;38(2):191-204. [doi: 10.1080/03069881003600991]
45. Mok K, Jorm AF, Pirkis J. Who goes online for suicide-related reasons? Crisis 2016 Mar;37(2):112-120. [doi: 10.1027/0227-5910/a000366] [Medline: 27232427]

46. Birnbaum ML, Candan K, Libby I, Pascucci O, Kane J. Impact of online resources and social media on help-seeking behaviour in youth with psychotic symptoms. Early Interv Psychiatry 2016 Oct;10(5):397-403. [doi: 10.1111/eip.12179] [Medline: 25200743]
47. Birnbaum ML, Rizvi AF, Confino J, Correll CU, Kane JM. Role of social media and the Internet in pathways to care for adolescents and young adults with psychotic disorders and non-psychotic mood disorders. Early Interv Psychiatry 2017 Aug;11(4):290-295 [FREE Full text] [doi: 10.1111/eip.12237] [Medline: 25808317]
48. Burns JM, Davenport TA, Durkin LA, Luscombe GM, Hickie IB. The internet as a setting for mental health service utilisation by young people. Med J Aust 2010 Jun 7;192(S11):S22-S26. [doi: 10.5694/j.1326-5377.2010.tb03688.x] [Medline: 20528703]
49. Best P, Manktelow R, Taylor B. Online communication, social media and adolescent wellbeing: a systematic narrative review. Child Youth Serv Rev 2014;41:27-36. [doi: 10.1016/j.childyouth.2014.03.001]
50. Barker SF, Manning V, Best DW, Savic M, Lubman DI. Alcohol, drug and related health and wellbeing issues among young people completing an online screen. Australas Psychiatry 2017 Apr;25(2):130-134. [doi: 10.1177/1039856216671657] [Medline: 27694645]
51. Burns J, Birrell E. Enhancing early engagement with mental health services by young people. Psychol Res Behav Manag 2014;7:303-312 [FREE Full text] [doi: 10.2147/PRBM.S49151] [Medline: 25473320]
52. Jorm AF, Korten AE, Jacomb PA, Christensen H, Rodgers B, Pollitt P. 'Mental health literacy': a survey of the public's ability to recognise mental disorders and their beliefs about the effectiveness of treatment. Med J Aust 1997 Feb 17;166(4):182-186. [Medline: 9066546]
53. Deci EL, Ryan RM. Self-determination theory: a macrotheory of human motivation, development, and health. Can Psychol 2008;49(3):182-185. [doi: 10.1037/a0012801]
54. Ryan R, Deci E. Self-determination Theory: Basic Psychological Needs In Motivation, Development, And Wellness. New York: The Guilford Press; 2017.
55. Peters D, Calvo RA, Ryan RM. Designing for motivation, engagement and wellbeing in digital experience. Front Psychol 2018;9:797 [FREE Full text] [doi: 10.3389/fpsyg.2018.00797] [Medline: 29892246]
56. Calvo R, Peters D. Positive Computing: Technology For Wellbeing And Human Potential. Cambridge, Massachusetts: The Mit Press; 2017.
57. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 2000 Jan;55(1):68-78. [doi: 10.1037//0003-066x.55.1.68] [Medline: 11392867]
58. Deci EL, Ryan RM. The 'what' and 'why' of goal pursuits: human needs and the self-determination of behavior. Psychol Inq 2000;11(4):227-268. [doi: 10.1207/S15327965PLI1104_01]

Abbreviations

CASP: Critical Appraisal Skills Program
SDT: self-determination theory

Edited by G Eysenbach; submitted 01.03.19; peer-reviewed by B Pang, A Barrett; comments to author 18.05.19; revised version received 22.07.19; accepted 19.08.19; published 19.11.19.

Please cite as:
Pretorius C, Chambers D, Coyle D
Young People's Online Help-Seeking and Mental Health Difficulties: Systematic Narrative Review
J Med Internet Res 2019;21(11):e13873
URL: http://www.jmir.org/2019/11/e13873/
doi:10.2196/13873
PMID:31742562

Original Paper

Using Cluster Analysis to Explore Engagement and e-Attainment as Emergent Behavior in Electronic Mental Health

Samineh Sanatkar[1], PhD; Peter Andrew Baldwin[1], PhD, MPsychol (Clin); Kit Huckvale[1], PhD; Janine Clarke[1], PhD, MPsychol (Clin); Helen Christensen[1], PhD; Samuel Harvey[1], MBBS, MRCGP, MRCPsych, FRANZCP, PhD; Judy Proudfoot[1], PhD

Black Dog Institute, School of Psychiatry, University of New South Wales, Sydney, Australia

Corresponding Author:
Samineh Sanatkar, PhD
Black Dog Institute
School of Psychiatry
University of New South Wales
Hospital Road
Sydney 2031,
Australia
Phone: 61 2 9382 ext 4368
Email: s.sanatkar@unsw.edu.au

Abstract

Background: In most e-mental health (eMH) research to date, adherence is defined according to a trial protocol. However, adherence to a study protocol may not completely capture a key aspect of why participants engage with eMH tools, namely, to achieve personal mental health goals. As a consequence, trial attrition reported as non-adherence or dropout may reflect *e-attainment*, the discontinuation of eMH engagement after personal goals have been met. Clarifying engagement patterns, such as e-attainment, and how these align with mental health trajectories, may help optimize eMH design and implementation science.

Objective: This study aimed to use clustering techniques to identify real-world engagement profiles in a community of eMH users and examine if such engagement profiles are associated with different mental health outcomes. The novelty of this approach was our attempt to identify actual user engagement behaviors, as opposed to employing engagement benchmarks derived from a trial protocol. The potential of this approach is to link naturalistic behaviors to beneficial mental health outcomes, which would be especially informative when designing eMH programs for the general public.

Methods: Between May 2013 and June 2018, Australian adults (N=43,631) signed up to myCompass, a self-guided eMH program designed to help alleviate mild to moderate symptoms of depression, anxiety, and stress. Recorded usage data included number of logins, frequency of mood tracking, number of started and completed learning activities, and number of tracking reminders set. A subset of users (n=168) completed optional self-assessment mental health questionnaires (Patient Health Questionnaire-9 item, PHQ-9; Generalized Anxiety Disorder Questionnaire-7 item, GAD-7) at registration and at 28 and 56 days after sign-up. Another subset of users (n=861) completed the PHQ-9 and GAD-7 at registration and at 28 days.

Results: Two-step cluster analyses revealed 3 distinct usage patterns across both subsamples: moderates, trackers, and super users, signifying differences both in the frequency of use as well as differences in preferences for program functionalities. For both subsamples, repeated measures analysis of variances showed significant decreases over time in PHQ-9 and GAD-7 scores. Time-by-cluster interactions, however, did not yield statistical significance in both subsamples, indicating that clusters did not predict symptom reduction over time. Interestingly, users who completed the self-assessment questionnaires twice had slightly but significantly lower depression and anxiety levels at sign-up compared with users who completed the questionnaires a third time at 56 days.

Conclusions: Findings suggested that although users engaged with myCompass in different but measurable ways, those different usage patterns evoked equivalent mental health benefits. Furthermore, the randomized controlled trial paradigm may unintentionally limit the scope of eMH engagement research by mislabeling early mental health goal achievers as dropouts. More detailed and naturalistic approaches to study engagement with eMH technologies may improve program design and, ultimately, program effectiveness.

(J Med Internet Res 2019;21(11):e14728) doi:10.2196/14728

KEYWORDS

eHealth; engagement; adherence; Web-based intervention; depression; anxiety

Introduction

Background

In 2005, Eysenbach [1] proposed a law of attrition to refer to the inevitability of substantial participant dropout (ie, dropout attrition) and discontinuation of program engagement (ie, nonusage attrition) from e-mental health (eMH) trials and called for rigorous examination of patterns of attrition to help clarify the impact, uptake, and dissemination of eMH programs. In reply, Christensen and Mackinnon [2] suggested that particular attention be given to users who derive symptom benefits from patterns of program engagement that deviate from trial protocols, particularly if nonusage attrition is common in randomized controlled trials (RCTs). Despite this, engagement in eMH research is still most frequently operationalized as adherence to an RCT protocol [3]. As such, much about program engagement outside the RCT context remains unknown. Where eMH programs are set up as a public health resource, understanding real-world engagement beyond protocol adherence is likely to strengthen eMH implementation science and help clarify patterns of eMH use in the general population. Findings of a recent systematic review of real-world engagement with eMH [4] suggest that attrition is also a common issue in naturalistic settings. Of the eleven interventions that were reviewed, 7% to 42% of sign-ups constituted moderate users who continually engaged with their eMH tool 4 to 6 weeks after registration. Only between 0.5% and 28.6% of users completed all modules or used the eMH program for more than 6 weeks after registration. As Price et al [5] pointed out, high dropout rates "limit the conclusions that can be drawn on the efficacy, feasibility, and public health impact of Web-based treatments."

Inferring user adoption of open-access eMH programs from participant engagement in an RCT can be problematic. eMH program delivery in the RCT context likely involves more comprehensive and structured user support than would be sustainable in a community setting [6]. Indeed, when eMH program use is compared across RCT and open-access conditions, dropout attrition appears to be delayed or reduced in the RCT environment. For example, Christensen et al [7] provided a comparison between questionnaire completion rates of Australians who signed up to MoodGYM, an internet-delivered mental health intervention, either via the program website or as part of an RCT. The authors found that 66.5% of RCT participants completed a depression measure at registration and at least one depression assessment at follow-up, whereas only 15.6% of general public MoodGYM users did so. Such findings may be due, in part, to follow-up by researchers seeking to maximize study adherence rates [1]. For these reasons, Cavanagh [6] suggested that extrapolating adherence criteria from RCTs and applying these to open-access interventions may be problematic, as RCT protocols are at odds with the design and intent of such programs. More recently, Sieverink et al [8] proposed that treating all eMH programs as linear courses of treatment may create a black box of eMH, where mediating mechanisms and nonlinear patterns of engagement go unexamined, simply because they are off protocol. Yardley et al [9] suggested that researchers give attention to elucidating patterns of effective engagement—engagement that leads to the desired benefits—in addition to more standard RCT-defined adherence measures.

Recent efforts to define adherence and engagement have suggested avenues for improving eMH engagement research. Sieverink et al [10] reviewed definitions of adherence across over 60 eMH trials and found that studies providing clear adherence criteria (eg, completing 75% of modules) often provided no theory or data to support these criteria, and most studies defined adherence as simply *the more use, the better*. These authors recommended that future definitions of adherence include multiple metrics of engagement, be data-driven, and reflect both the goals of the technology (eg, symptom reduction) and the goals of the individual (eg, subjective states of well-being). In their recent scoping review, Pham et al [11] concluded that, although engagement metrics are increasingly consistent across studies, it is still unclear how patterns of engagement relate to mental health outcomes. Furthermore, Ryan et al [3] identified 8 theoretical frameworks in which adherence could be defined but found that few studies reported comprehensive program adherence data, with almost no consistency in the frameworks used to define engagement, attrition, and adherence. In line with Eysenbach [1] and Christensen and Mackinnon [2], Ryan et al [3] highlighted the ongoing risk of conflating engagement with adherence and the need to extend our understanding of eMH engagement beyond efficacy trials.

Eysenbach [1] proposed that a potential way forward was to examine the shape of attrition curves, a procedure that could hint at the presence of behavioral groupings, from low users to hardcore users. This suggests that one method for studying effective program engagement may be to identify *engagement profiles* within eMH cohorts and to examine whether program benefits (if any) vary between them. Some efforts have already been made to identify such usage profiles within RCTs. For instance, Donkin et al [12] examined which behaviors predicted clinically significant reductions in depressive symptoms among adherent users of a Web-based intervention targeting depression in cardiovascular disease. Activities per user session predicted significant improvement, whereas time spent on the intervention or the total number of modules and activities completed did not. High in-session engagers reported the best outcomes; however, medium and low in-session engagers experienced equivalent benefits. These results suggest that even within an adherent group the dose-response relationship is nonlinear and in-session engagement behavior may be more important to program efficacy than overall time-on-task. On the basis of their findings, Donkin et al [12] suggest that identifying when and for whom *treatment saturation* is achieved is an important goal for future eMH engagement research.

Techniques common in other areas of cognitive behavioral science may further help to identify eMH engagement profiles.

Clustering and classification techniques can help identify emergent patterns of behavior [13]. Using engagement metrics that correspond to specific on-task behaviors, individuals can be grouped based on how they use their chosen eMH program. Once grouped, comparisons between these classes or clusters can be made to determine if different engagement profiles experience different benefits. The advantage of such an approach is that user groups emerge from the usage data. The results of such research are likely to better reflect user adoption behavior in an open-access environment with important implications for eMH implementation science.

Objectives

The purpose of this study was to explore the utility of using clustering approaches to identify engagement profiles in eMH and to examine if such engagement profiles lead to different mental health outcomes. This approach has the potential to identify actual user engagement behaviors that may result in beneficial mental health outcomes and is likely to be especially informative when designing eMH programs for the general public. Cluster analytical approaches can help identify which short-term users, or dropouts, may be classified as e-attainers (ie, those who discontinue using an eMH program when ones' personal mental health objectives have been attained, see [2]) and conversely, which sustained users may not achieve their desired mental health benefits. By understanding the link between usage behavior and mental health symptom progression, patterns of eMH tool cessation allow for a more differentiated view of attrition.

Methods

Target Program

myCompass is a self-guided Web-based mental health program available free to Australians [14]. It is designed for individuals experiencing mild to moderate symptoms of depression, anxiety, and stress and offers symptom tracking and interactive learning activities based on cognitive behavioral therapies. The "Daily Tracker" feature allows real-time tracking (by mobile phone or computer) of up to 3 areas of difficulty that may be emotional (eg, depression or anxiety), cognitive (eg, worry), or behavioral (eg, smoking) by measuring each on an 11-point scale ranging from low (0) to high (10). Users can also set to receive tracking reminders via SMS or email.

The myCompass learning activities aim to teach techniques for managing distress and improving well-being. Learning activities are delivered in 3 sessions and contain a session introduction, didactic content, interactive activities, and a practical home task. Each learning activity takes approximately 10 to 15 min to complete. myCompass also prompts users to complete an optional "self-assessment" questionnaire at 3 time points: the time of registration and at 28 and 56 days after sign-up. The self-assessment questionnaire (described below) measures symptoms of anxiety and depression and takes about 4 min to complete.

Sample

The data were extracted from the usage data obtained from 43,631 adults who signed up to myCompass between May 2013 and June 2018. Upon signing up and as part of the Terms of Use agreement, all myCompass users consented to have their deidentified demographic and program usage information used for research purposes.

We conducted 2 sample extractions. In our first sample extraction, users were eligible for inclusion in analyses if they had completed self-assessment questionnaire data for all 3 time points (n=168). In our second sample extraction, we broadened our eligibility criteria to include individuals who completed the self-assessment questionnaire at the first 2 time points only (n=861 additional individuals). The resulting 1029 myCompass users included in this analysis had a mean age of 40.9 years (SD 13.45) and were mostly female (63.07%, 649/1029). Most resided in the Australian state of New South Wales (44.31%, 456/1029), followed by Victoria (19.63%, 202/1029) and Queensland (12.83%, 132/1029). Overall depression and anxiety scores, measured using the 9-item Patient Health Questionnaire (PHQ-9) [14] and the 7-item General Anxiety Disorder (GAD-7) Scale [15], were in the mild symptom ranges at sign-up (mean 10.68, SD 6.46 for depression and mean 8.58, SD 5.43 for anxiety).

Engagement Metrics and Self-Assessment Questionnaire

Engagement Metrics

We used 5 usage metrics in our cluster analyses: number of user logins, number of daily trackers used, number of learning activities started, number of learning activities completed, and number of reminders received. All engagement data were extracted from and stored on secure myCompass servers, which record all user activity across the myCompass website. We took a conservative approach and defined login and tracking attempts that occurred in close succession (eg, 30 min apart for logins and less than 24 hours apart for daily tracking) as representing relogins because of system time-out. Hence, only the first instance of such records was included in our analyses.

Self-Assessment Questionnaire

The self-assessment questionnaire comprises the PHQ-9 and the GAD-7 and is used to measure symptoms of depression and generalized anxiety disorder, respectively. Both questionnaires ask individuals to rate symptom severity over the last 2 weeks on a 4-point Likert-type scale anchored *Not at all* (0) and *Nearly every day* (3). An example item for the PHQ-9 is "Little interest or pleasure in doing things," and an example item for the GAD-7 is "Feeling nervous, anxious or on the edge." Values on the PHQ-9 range from 0 to 27, and values on the GAD-7 range from 0 to 21. Both scales use cutoff scores of 5, 10, and 15 to identify individuals with mild, moderate, and moderately severe depressive and anxiety symptoms, respectively. The scales have been used extensively in clinical and research domains.

Analysis Strategy

All analyses were computed in SPSS v.25 (IBM Corp). First, we employed a 2-step cluster analysis procedure to group users based on similarities in their myCompass engagement metrics. Cluster analysis is an exploratory classification technique that statistically identifies groupings in a dataset based on the degree

of similarity between specified data points (for a more thorough discussion, see Jain et al [13]). A 2-step cluster analytical approach first constructs a cluster feature tree based on the similarity between data points and then uses a clustering algorithm to group the cluster feature tree nodes into an optimal number of clusters. This clustering procedure handles large datasets in a time-efficient manner and is quite robust against violations of assumptions.

As mentioned, the first cluster analysis contained only those 168 users who provided PHQ-9 and GAD-7 data at all 3 time points: sign-up, 28 days, and 56 days. To increase power and determine if our initial clusters would hold across a larger group, the second cluster analysis was broadened to contain users who provided mental health data at sign-up and at 28 days (but not at 56 days). In both analyses, we selected the loglikelihood distance measure and used the Bayesian information criterion as the clustering criterion. We let the program automatically compute the optimal number of clusters and used the default option of allowing no more than 15 clusters to be selected. To verify that the procedure had generated meaningfully different clusters, we used multivariate analysis of variance (ANOVA) to compare the clusters on all engagement metrics. We then employed repeated-measures ANOVA to examine if PHQ-9 and GAD-7 scores changed across time and differed between clusters.

Results

Analysis 1

Two-Step Cluster Analysis

The cluster analysis identified 3 distinct clusters (see Table 1). The first cluster, which we labeled “moderates,” comprised 74 users who, on average, logged into myCompass 11 times, used the daily tracker about 7 times, and completed 1 learning activity during 2 months of using myCompass. The second cluster, labeled “trackers,” contained 69 users who accessed myCompass about 36 times, used the daily tracker an average of 34 times, but showed comparable learning activity completion rates compared with the previous cluster across the 2-month time span. The third cluster “super users” represented 25 users who logged in approximately 39 times, used the daily tracker an average of 35.28 times, and showed markedly higher learning activity completion than the other 2 clusters (mean=5 modules completed). Multivariate ANOVA confirmed that the 3 cluster groups were significantly different on all usage variables (*P* values<.001), indicating that the cluster analysis identified distinct user groups.

Table 1. Results from multivariate analysis of variance looking at usage behaviors based on cluster membership of 168 myCompass users who completed the symptom screener at sign-up, 28 days, and 56 days.

Usage variables[a] (count data)	Moderates (n=74), mean (SE)	Trackers (n=69), mean (SE)	Super users (n=25), mean (SE)	*F* test (*df*)	*P* value
Logins	11.00 (1.09)[b]	36.26 (1.13)[c]	38.92 (1.87)[c]	160.69 (2,165)	<.001
Tracking	6.91 (1.16)[b]	34.45 (1.21)[c]	35.28 (2.00)[c]	159.48 (2,165)	<.001
Modules completed	0.97 (0.13)[b]	0.80 (0.13)[b]	4.48 (0.22)[c]	116.41 (2,165)	<.001
Modules started	1.99 (0.15)[b]	1.33 (0.15)[c]	5.36 (0.26)[d]	93.10 (2,165)	<.001
Notifications received	5.53 (0.65)[b]	13.43 (0.67)[c]	9.80 (1.11)[d]	36.21 (2,165)	<.001

[a]Input variables are sorted by overall importance in this analysis, from highest to lowest.

[b,c,d]Differing superscripts indicate a significant difference at $P<.05$ for Bonferroni corrected multiple comparisons.

Repeated-Measures Analyses

The Mauchly test of sphericity indicated that the variances of differences between groups were not equal. Hence, we proceeded to use Greenhouse-Geisser corrections to adjust the degrees of freedom. As shown in Table 2 and Figures 1 and 2, all clusters reported significant reductions in PHQ-9 and GAD-7 scores over time. However, the time-by-cluster interactions did not reach significance, indicating that symptom reduction over time did not vary by cluster in this initial sample.

Table 2. Repeated measures analysis of variances on Patient Health Questionnaire-9 item and Generalized Anxiety Disorder Scale-7 item scores at sign-up, 28 days, and 56 days compared across clusters (N=168).

Mental health variables	*F* test (*df*)	*P* value
PHQ-9[a]		
Time	30.38 (1.87,309.23)	<.001
Time × clusters	2.01 (3.75,309.23)	.10
GAD-7[b]		
Time	23.84 (1.76,287.07)	<.001
Time × clusters	1.05 (3.52,287.07)	.38

[a]PHQ-9: Patient Health Questionnaire-9 item.

[b]GAD-7: Generalized Anxiety Disorder Scale-7 item.

Figure 1. Mean trends of depression symptom progression, as measured by the Patient Health Questionnaire-9 item (range: 0-27), for myCompass usage groups at sign-up, 28 days, and 56 days.

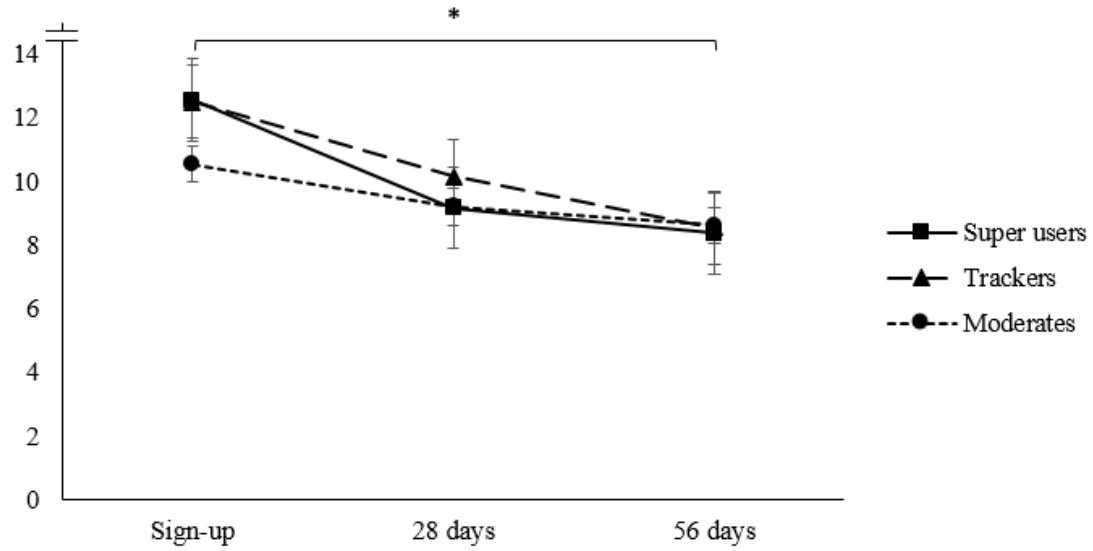
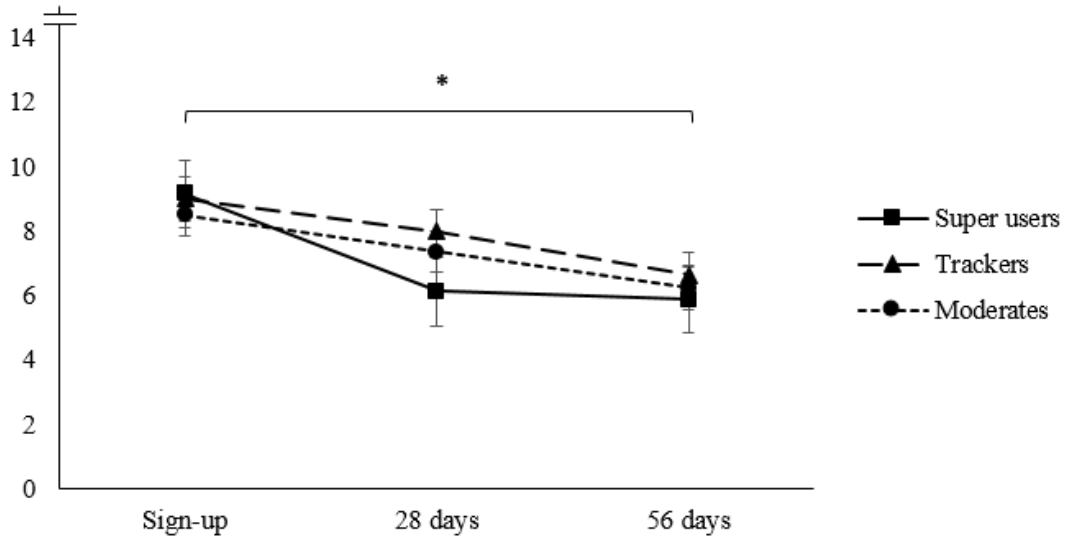


Figure 2. Mean trends of anxiety symptom progression, as measured by the Generalized Anxiety Disorder Scale-7 item (range: 0-21), for myCompass usage groups at sign-up, 28 days, and 56 days.



Analysis 2

Two-Step Cluster Analysis

We repeated the steps of the first cluster analysis on the sample of 861 unique users who completed the PHQ-9 and GAD-7 at sign-up and at 28 days. As in the previous sample, the 2-step cluster analysis yielded 3 distinct user groups (see Table 3) that largely fit the moderates, trackers, and super user clusters previously identified. Multivariate ANOVA analysis confirmed that the newly derived 3 clusters also differed significantly on all usage variables (*P* values<.001).

Repeated-Measures Analyses

Repeated measures ANOVAs showed significant decreases over time in PHQ-9 and GAD-7 scores; however, the time-by-cluster interactions were not significant (see Table 4 and Figures 3 and 4).

Table 3. Results from multivariate analysis of variance looking at usage behaviors based on cluster membership of 861 myCompass users who completed the symptom screener at sign-up and at 28 days.

Usage variables[a] (count data)	Moderates (n=479), mean (SE)	Trackers (n=224), mean (SE)	Super users (n=158), mean (SE)	*F* test (*df*)	*P* value
Modules completed	0.18 (0.48)[b]	0.23 (0.40)[b]	2.08 (0.62)[c]	651.68 (2,858)	<.001
Logins	4.92 (0.33)[b]	19.02 (0.27)[c]	13.25 (0.43)[d]	960.89 (2,858)	<.001
Tracking	3.44 (0.34)[b]	17.96 (0.28)[c]	10.44 (0.44)[d]	930.79 (2,858)	<.001
Modules started	0.85 (0.07)[b]	0.77 (0.06)[b]	3.16 (0.09)[c]	443.56 (2,858)	<.001
Notifications received	5.22 (0.31)[b]	9.05 (0.25)[c]	6.28 (0.39)[d]	78.79 (2,858)	<.001

[a]Input variables are sorted by overall importance in this analysis, from highest to lowest.

[b,c,d]Differing superscripts indicate a significant difference at *P*<.05 for Bonferroni corrected multiple comparisons.

Table 4. Repeated measures analysis of variance on Patient Health Questionnaire-9 item and Generalized Anxiety Disorder Scale-7 item scores at sign-up and at 28 days compared across clusters.

Mental health variables.	*F* test (*df*)	*P* value
PHQ-9[a]		
Time	66.99 (1,858)	<.001
Time × clusters	0.67 (2,858)	.51
GAD-7[b]		
Time	42.99 (1,858)	<.001
Time × clusters	0.83 (2,858)	.44

[a]PHQ-9: Patient Health Questionnaire-9 item.

[b]GAD-7: Generalized Anxiety Disorder Scale-7 item, N=861.

Figure 3. Mean trends of depression symptom progression, as measured by the Patient Health Questionnaire (range = 0 - 27), for myCompass usage groups at sign-up and after 28 days.

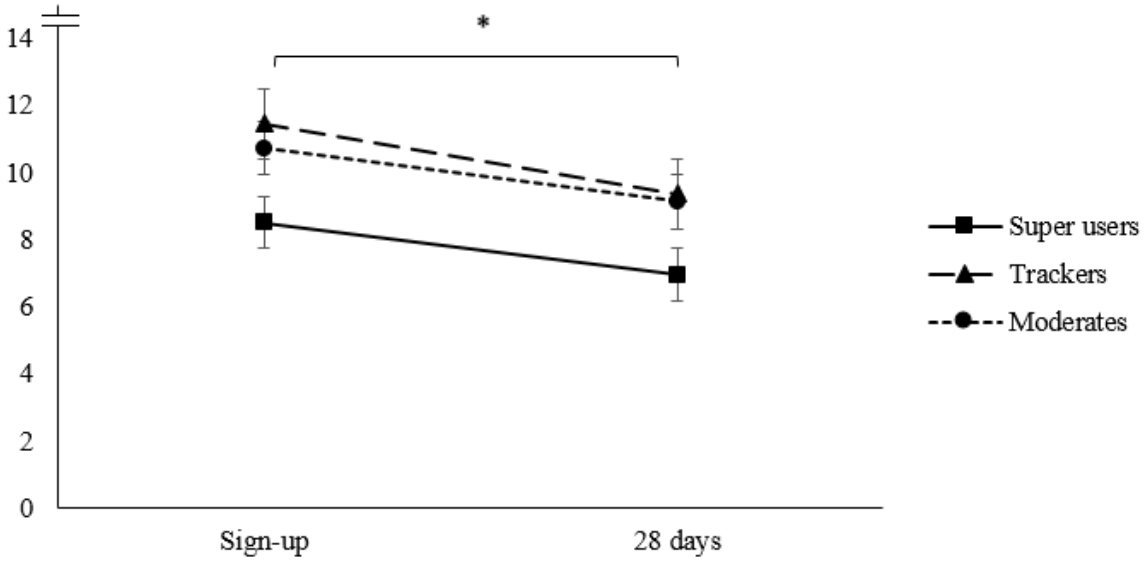
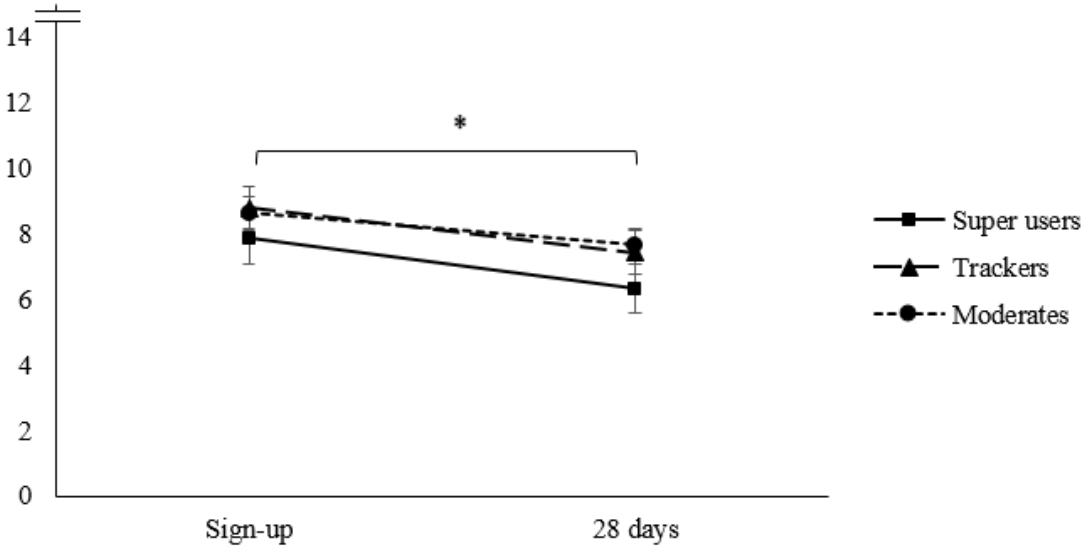


Figure 4. Mean trends of anxiety symptom progression, as measured by the Generalized Anxiety Disorder Scale-7 item (range: 0-21), for myCompass usage groups at sign-up and after 28 days.



Subsample Comparisons

In the second analysis, the super users reported the *lowest* PHQ-9 and GAD-7 scores at sign-up, contrasting with the symptom pattern observed in the first analysis. To determine if this effect was because of the inclusion of participants who completed self-report measures less frequently, we used ANOVA analyses to compare the 2 subsamples on their characteristics at the time of registration. Users who completed the self-assessment questionnaire 3 times were older, had slightly higher baseline PHQ-9 and GAD-7 scores, and used myCompass more in the first 28 days after signing up (see Table 5) compared with users who only completed the self-assessment questionnaire twice. Table 5 further shows differences between the characteristics of the 2 subsamples and the overall sample. Usage across the first 28 days after registration was significantly higher for myCompass users who revisited the platform over 1 or 2 consecutive months compared with the full sample that contained a large number of 1-time users (23,688/43,631, 54.29%).

Table 5. Means, SDs, and between-group statistics on demographic information and symptom severity at sign-up and usage behavior across 28 days.

Baseline characteristics and 28-day usage	Full sample (N=43,631), mean (SD)	2-month subsample (N=168), mean (SD)	1-month subsample (N=861), mean (SD)	*F* test	*P* value
Age	39.09[a] (13.35)	41.29[b] (14.14)	40.83[c] (13.31)	9.37 (2,44657)	<.001
Females	1.69[a] (0.47)	1.65[b] (0.48)	1.63[b] (0.49)	7.44 (2,44657)	.001
Baseline PHQ-9[d]	11.28[a] (6.69)	11.64[a] (6.53)	10.50[b] (6.44)	5.99 (2,44657)	<.001
Baseline GAD-7[e]	8.98[a] (5.73)	8.89[a] (5.36)	8.52[b] (5.45)	2.73 (2,44657)	.07
Logins	3.18[a] (4.93)	14.64[b] (8.82)	10.12[c] (7.36)	1238.73 (2,44657)	<.001
Tracking	1.74[a] (3.82)	12.75[b] (9.49)	8.51[c] (7.48)	1866.71 (2,44657)	<.001
Modules started	0.47[a] (0.83)	1.64[b] (1.39)	1.25[c] (1.27)	517.99 (2,44657)	<.001
Modules completed	0.10[a] (.47)	0.81[b] (1.08)	0.54[c] (.94)	518.18 (2,44657)	<.001
Notifications	3.19[a] (3.33)	6.88[b] (4.19)	6.41[b] (4.11)	488.07 (2,44657)	<.001

[a,b,c]Differing superscripts indicate significant differences of $P<.05$ between groups.

[d]PHQ-9: Patient Health Questionnaire-9 item.

[e]GAD-7: Generalized Anxiety Disorder Scale-7 item.

Discussion

Principal Findings

The aim of this study was to investigate the feasibility and utility of using a clustering procedure to group eMH users based on their engagement behavior, and to determine if these engagement clusters reported any difference in mental health outcomes. Using a 2-step clustering procedure, we identified 3 clusters of engagement behavior, which emerged consistently across 2 subsamples in our community dataset. Although some patterns in the data suggest that increased overall engagement could be linked with more rapid symptom improvement, all users experienced equivalent symptomatic relief over the course of program use, irrespective of their pattern of engagement, which may speak to e-attainment.

The time-by-group differences were not statistically significant in analysis 1. As the study used naturalistic data that had already been collected—akin to archival data, we were unable to conduct an a priori power analysis to set a target sample size. Hence, it is possible that some of the statistical tests were underpowered and warrant replication in comparable but larger datasets. Therefore, any interpretations of the findings as they are presented in this study need to be made with caution. Bearing these considerations in mind, we would like to discuss interesting patterns that emerged in the data.

In analysis 1, “moderates” started out with slightly (though not significantly) lower depression scores at sign-up compared with “trackers” and “super users.” At 28 days, moderates and super users reported only mild depressive symptoms, whereas trackers remained just within the moderate symptom range. Only super users reported a clinically significant reduction in depressive symptoms (>3 points) at 28 days, though this likely reflects their higher PHQ-9 scores at the commencement of their program use. Notably, by 56 days all groups reported that their depressive symptoms had reduced into the mild symptom range.

Similar nonsignificant patterns emerged for anxiety symptoms in analysis 1. Moderates reported slightly lower anxiety levels at sign-up compared with trackers and super users. Super users showed the greatest reduction in anxiety scores at 28 days compared with trackers and moderates. At 56 days, all user groups had progressed further to the lower end of the mild anxiety range.

Consistent with pattern of results from analysis 1, super users in analysis 2 reported significantly lower depression scores at 28 days compared with moderates and trackers, though all groups reported depressive symptoms in the mild range by 28 days. Users’ anxiety symptom progression in analysis 2 also resembled the pattern of results for depression. At registration, average anxiety scores were in the mild anxiety range. At 28 days, all users progressed further toward the mid and bottom values of the mild anxiety category, with super users reporting significantly lower levels of anxiety than both moderates and trackers.

There were some interesting differences between our super users across the 2 analyses. In analysis 1 (across 2 months), super users appeared to report the *highest* distress at sign-up. However, in analysis 2 (across 1 month), super users appeared to report the *lowest* distress at sign-up compared with moderates and trackers. More generally, we found that the 1-month users from analysis 2 reported mild but significantly lower distress levels

at sign-up compared with the more sustained 2-month users from analysis 1. It may be the case that some users experiencing greater distress at sign-up continued to use myCompass for longer than other users with milder symptoms.

Interestingly, belonging to a cluster in our study did not always seem to reflect symptom severity alone. Therefore, these engagement profiles may reflect both user need at sign-up and individual difference factors in technology engagement. For example, the super user engagement profile was identical across both analyses, but symptom severity at sign-up was different for 1-month super users relative to 2-month super users. This speaks to the complexity of eMH engagement behavior and underscores recommendations made by previous researchers [9,10] that individual user characteristics be included in conceptualizations of eMH engagement. If individuals ceased use of myCompass according to their symptom levels, we may have ended up with a less symptomatic group in our 1-month sample, relative to our 2-month sample. This systematic effect on program engagement would likely have resulted in symptom profile disparities between our analyses. Therefore, these seemingly contradictory results may speak to the phenomenon of e-attainment suggested by Christensen and Mackinnon [2], whereby users engage with an eMH tool as much and for as long as they need to reach their mental health goals. An alternative interpretation for the results would be that some of the standard components of eMH programs do not have the expected impact on symptom levels. For example, our results showed that trackers appeared to do as well as super users, which gives the impression that learning activities do not provide additional benefit beyond the effects of mood monitoring for some users. Given that substantial empirical evidence supports the cognitive behavioral therapies used in the learning activities, more data are required to better understand the unique benefits of specific eMH treatment components.

Our study provided an attempt at opening the black box of eMH [8] by moving beyond a priori definitions of adherence and, instead, inspecting behavior as it emerged from open-access program usage. Our findings suggested that, at least among our self-selected sample, individuals seem to engage with our eMH program in different but measurable ways that lead to equivalent mental health benefits. These findings align with a previous investigation conducted by Matthews et al [15]. Similar to the current investigation, the authors used cluster analysis to inspect naturalistic usage patterns among mobile phone users who downloaded and registered to a self-help app designed to help reduce anxiety symptoms. Findings indicated that all 4 usage groups identified experienced short-term decreases of anxiety symptoms irrespective of their engagement profile. Where mental health goals are achieved in a relatively short period of time, cessation of program use may create the impression of dropout across eMH websites and mobile phone apps.

Whereas the robustness of the RCT paradigm is crucial to efficacy science [16], it may unintentionally limit the scope of eMH engagement research in certain ways. For example, statistical practices that assume no benefit for participants who deviate from the trial protocol (eg, per-protocol or intention-to-treat with last-operation-carried forward analyses) may exclude individuals for whom program use is nonetheless beneficial. In addition, the prevailing RCT paradigms of symptom reduction or diagnostic remission may not reflect the usage goals or expectations of all eMH users. Although some eMH consumers seek treatment for a specific mental health condition, others may use eMH programs to navigate periods of general distress or to build resilience [17,18]. Many eMH users may be capable of determining when sufficient help has been accessed and subsequently discontinue without ever measuring symptoms or receiving a diagnosis [2]. However, within a traditional RCT framework, these so-called e-attainers are likely to be classified as nonadherent and therefore excluded from further analysis or aggregated with true nonusers in any subsequent per protocol analyses. Traditional RCT designs may be able to follow up participants who ceased using their prescribed eMH program and get an indication as to why participants opted out of the program. This approach was successfully implemented by Postel et al [19] who followed up RCT participants who prematurely ended their involvement with a Web-based intervention to reduce problem drinking. Interestingly, satisfaction with the treatment progress constituted the third most common reason for participants' decision to drop out of the study, highlighting the potential importance of identifying whether noncontinuation is because of intervention success or failure. Although findings are preliminary, our results support Yardley et al's [9] recommendations for identifying effective engagement and Pham et al's [11] appeal to move beyond descriptive reporting of usage behaviors and examine the relationship between engagement and mental health outcomes. These efforts could provide a valuable supplement to traditional a priori definitions of eMH engagement.

Limitations

Some limitations of our study must be acknowledged. First, the symptom change data were uncontrolled and therefore cannot be compared against a comparison condition as we inspected usage behavior in a naturalistic setting. Consequently, symptom reduction over time may be because of remission patterns rather than reflective of a treatment effect. Second, it is important to note that our clustering procedure is only one example of a categorization procedure, and other methods, such as latent class analysis, may yield different results. As mentioned above, discussing trends can be useful but further research is required to see if these effects can reach significance and are replicable in other datasets. The purpose of this study was not to establish best practice, rather it was to describe a novel way of inspecting usage behavior in the hope of encouraging similar approaches in the future that examine naturalistic behavior in addition to predetermined behavioral benchmarks. Finally, it is important to note that entry into the sample used for this study was dependent on users completing the symptom assessment scales at least twice. Therefore, there is at least one other group of users of eMH products not included in this analysis, those who do not engage at all following the initial assessment or who engage very little. The impact of this pattern of nonengagement remains unclear.

Conclusions

The benefits of exploring real-world examples of engagement is to gain a more differentiated picture of how users navigate

through the eMH space and, by doing so, to advance our understanding of how eMH tools might become more sophisticated and helpful companions in mental health. Learning from implicit engagement patterns can help inform and strengthen new computational techniques, such as machine learning, which aim to provide a personalized and situation-sensitive user experience in real time and are designed to motivate users to engage with their respective eMH tool until the desired mental health goals are reached and sustained.

Examination of real-world eMH engagement is required to assist differentiation between e-attainment and dropout and will have important implications for progressing the eMH space in a way that results in more widespread acceptance of eMH tools in the general public. Ultimately, a primary goal of eMH research is to remove obstacles to engagement for individuals who might otherwise benefit from eMH. A more nuanced view of how many variables such as limited access to the internet, a lack of fit between the needs and wants of users and the eMH program, and the achievement of individual goals (ie, e-attainment) will inform how future eMH technologies can better reflect the behaviors and desires of their users.

Conflicts of Interest

None declared.

References

1. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
2. Christensen H, Mackinnon A. The law of attrition revisited. J Med Internet Res 2006 Sep 29;8(3):e20; author reply e21 [FREE Full text] [doi: 10.2196/jmir.8.3.e20] [Medline: 17032636]
3. Ryan C, Bergin M, Wells JS. Theoretical perspectives of adherence to web-based interventions: a scoping review. Int J Behav Med 2018 Feb;25(1):17-29. [doi: 10.1007/s12529-017-9678-8] [Medline: 28730402]
4. Fleming T, Bavin L, Lucassen M, Stasiak K, Hopkins S, Merry S. Beyond the trial: systematic review of real-world uptake and engagement with digital self-help interventions for depression, low mood, or anxiety. J Med Internet Res 2018 Jun 6;20(6):e199 [FREE Full text] [doi: 10.2196/jmir.9275] [Medline: 29875089]
5. Price M, Gros DF, McCauley JL, Gros KS, Ruggiero KJ. Nonuse and dropout attrition for a web-based mental health intervention delivered in a post-disaster context. Psychiatry 2012;75(3):267-284 [FREE Full text] [doi: 10.1521/psyc.2012.75.3.267] [Medline: 22913502]
6. Cavanagh K. Turn on, tune in, (don't) drop out: Engagement, adherence, attrition, and alliance with internet-based interventions. In: Bennett-Levy J, Richards DA, Farrand P, Christensen H, Griffiths KM, Kavanaugh DJ, et al, editors. Oxford Guide to Low-Intensity CBT Interventions. New York: Oxford University Press; 2010:227-233.
7. Christensen H, Griffiths KM, Korten AE, Brittliffe K, Groves C. A comparison of changes in anxiety and depression symptoms of spontaneous users and trial participants of a cognitive behavior therapy website. J Med Internet Res 2004 Dec 22;6(4):e46 [FREE Full text] [doi: 10.2196/jmir.6.4.e46] [Medline: 15631970]
8. Sieverink F, Kelders S, Poel M, van Gemert-Pijnen L. Opening the black box of electronic health: collecting, analyzing, and interpreting log data. JMIR Res Protoc 2017 Aug 7;6(8):e156 [FREE Full text] [doi: 10.2196/resprot.6452] [Medline: 28784592]
9. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
10. Sieverink F, Kelders SM, van Gemert-Pijnen JE. Clarifying the concept of adherence to eHealth technology: systematic review on when usage becomes adherence. J Med Internet Res 2017 Dec 6;19(12):e402 [FREE Full text] [doi: 10.2196/jmir.8578] [Medline: 29212630]
11. Pham Q, Graham G, Carrion C, Morita PP, Seto E, Stinson JN, et al. A library of analytic indicators to evaluate effective engagement with consumer mHealth apps for chronic conditions: scoping review. JMIR Mhealth Uhealth 2019 Jan 18;7(1):e11941 [FREE Full text] [doi: 10.2196/11941] [Medline: 30664463]
12. Donkin L, Hickie IB, Christensen H, Naismith SL, Neal B, Cockayne NL, et al. Rethinking the dose-response relationship between usage and outcome in an online intervention for depression: randomized controlled trial. J Med Internet Res 2013 Oct 17;15(10):e231 [FREE Full text] [doi: 10.2196/jmir.2771] [Medline: 24135213]
13. Jain AK, Murty MN, Flynn PJ. Data clustering: a review. ACM Comput Surv 1999;31(3):264-323. [doi: 10.1145/331499.331504]
14. Proudfoot J, Clarke J, Birch M, Whitton AE, Parker G, Manicavasagar V, et al. Impact of a mobile phone and web program on symptom and functional outcomes for people with mild-to-moderate depression, anxiety and stress: a randomised controlled trial. BMC Psychiatry 2013 Nov 18;13:312 [FREE Full text] [doi: 10.1186/1471-244X-13-312] [Medline: 24237617]

15. Matthews P, Topham P, Caleb-Solly P. Interaction and Engagement with an Anxiety Management App: Analysis Using Large-Scale Behavioral Data. JMIR Ment Health 2018 Sep 14;5(4):e58 [FREE Full text] [doi: 10.2196/mental.9235] [Medline: 30287415]
16. Eysenbach G. The law of attrition revisited – author's reply. J Med Internet Res 2006;8(3):e21. [doi: 10.2196/jmir.8.3.e21]
17. Burns JM, Davenport TA, Durkin LA, Luscombe GM, Hickie IB. The internet as a setting for mental health service utilisation by young people. Med J Aust 2010;192(11 Suppl):S22-S26. [doi: 10.5694/j.1326-5377.2010.tb03688.x]
18. Klein B, Cook S. Preferences for e-mental health services amongst an online Australian sample? E J Appl Psychol 2010;6(1):28-39. [doi: 10.7790/ejap.v6i1.184]
19. Postel MG, de Haan HA, ter Huurne ED, van der Palen J, Becker ES, de Jong CA. Attrition in web-based treatment for problem drinkers. J Med Internet Res 2011 Dec 27;13(4):e117 [FREE Full text] [doi: 10.2196/jmir.1811] [Medline: 22201703]

Abbreviations

ANOVA: analysis of variance
eMH: e-mental health
GAD-7: generalized anxiety disorder questionnaire-7 item
PHQ-9: Patient Health Questionnaire-9 item
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 20.05.19; peer-reviewed by K Cavanagh, P Matthews; comments to author 29.06.19; revised version received 22.08.19; accepted 24.09.19; published 28.11.19.

Please cite as:
Sanatkar S, Baldwin PA, Huckvale K, Clarke J, Christensen H, Harvey S, Proudfoot J
Using Cluster Analysis to Explore Engagement and e-Attainment as Emergent Behavior in Electronic Mental Health
J Med Internet Res 2019;21(11):e14728
URL: http://www.jmir.org/2019/11/e14728/
doi:10.2196/14728
PMID:31778115

Original Paper

Associations Between Characteristics of Web-Based Diabetes News and Readers' Sentiments: Observational Study in the Netherlands

Hans Vehof[1,2*], MSc; Eibert Heerdink[2,3*], PhD; José Sanders[1*], PhD; Enny Das[1*], PhD

[1]Centre for Language Studies, Radboud University, Nijmegen, Netherlands

[2]Research Group Process Innovations in Pharmaceutical Care, HU University of Applied Sciences, Utrecht, Netherlands

[3]Division Pharmacoepidemiology & Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences, Utrecht, Netherlands

[*]all authors contributed equally

Corresponding Author:
Hans Vehof, MSc
Research Group Process Innovations in Pharmaceutical Care
HU University of Applied Sciences
Heidelberglaan 7
Utrecht, 3584 CS
Netherlands
Phone: 31 625098999
Email: hans.vehof@hu.nl

Abstract

Background: Although experts agree that Web-based health information often contains exaggeration and misrepresentation of science, it is not yet known how this information affects the readers' sentiments.

Objective: This study aimed to investigate whether specific aspects of Web-based diabetes research news are associated with positive or negative sentiments in readers.

Methods: A retrospective observational study of the comments on diabetes research news posted on Facebook pages was conducted as a function of the innovations' developmental phase, the intended treatment effect, and the use of strong language to intensify the news messages (superlatives). Data for the investigation were drawn from the diabetes research news posted between January 2014 and January 2018 on the two largest Dutch Facebook pages on diabetes and the corresponding reader comments. By manually coding these Facebook user comments, three binary outcome variables were created, reflecting the presence of a positive sentiment, the presence of a negative sentiment, and the presence of a statement expressing hopefulness.

Results: Facebook users made a total of 3710 comments on 173 diabetes research news posts that were eligible for further analysis. Facebook user comments on posts about diabetes prevention (odds ratio [OR] 0.55, 95% CI 0.37-0.84), improved blood glucose regulation (OR 0.68, 95% CI 0.56-0.84), and symptom relief (OR 0.31, 95% CI 0.21-0.44) were associated with less positive sentiments as compared with potential diabetes cures. Furthermore, comments on innovations supported by preclinical evidence in animals were associated with more positive sentiments (OR 1.46, 95% CI 1.07-1.99) and statements expressing hope (OR 1.47, 95% CI 1.01-2.14), when compared with innovations that have evidence from large human trials. This study found no evidence for the associations between language intensification of the news posts and the readers' sentiments.

Conclusions: Our finding that the attitudes toward diabetes research news on Facebook are most positive when clinical efficacy is not (or not yet) proven in large patient trials suggests that news authors and editors, as well as medical professionals, must exercise caution when acting as a conduit for diabetes research news.

(J Med Internet Res 2019;21(11):e14554) doi:10.2196/14554

KEYWORDS

medical journalism; diabetes mellitus; information seeking behaviors; news; diffusion of innovation

Introduction

Background

Patients who monitor online media for health information may experience frequent exposure to exaggeration and misrepresentation of medical science [1-5]. Two typical examples of such infelicitous reporting are the depiction of observed correlations as causal connections—for example, between lifestyle behaviors and disease outcomes—and the inflation of preclinical animal testing results, often followed by the inference of these to humans [6,7]. This is misleading when one considers that about 88% of the pharmaceutical developments that reach the first human trials will never reach the phase of market approval [8]. Such misrepresentations are present in numerous easily accessible health news sites and are spread freely on social media such as Facebook.

Earlier research by our group, on the media coverage of innovative diabetes therapies, found that 83% of Dutch newspaper reports about innovative diabetes treatments lack any reference to clinical trials in humans [9]. Similarly, in the United States, a study on health news appraisals found that most authors do not satisfactorily discuss the quality of the evidence [6].

Although, to our knowledge, there is no literature on the effects of news reporting on the patients' attitudes, it is highly plausible that messages about promising future treatments could affect the readers' sentiments such as enthusiasm and curiosity. Moreover, a patient's level of hope may increase, which is positive as having hope is associated with more favorable diabetes outcomes [10,11]. The effects may also turn out to be negative when, for example, feelings of impatience or disbelief are more prominent.

Overall, 3 aspects of Web-based reporting may influence attitudes. First, the tone of the reports, using intensified language (eg, *revolutionary* and *breakthrough*), a common phenomenon in health news coverage [12], may amplify these sentiments.

Second, attitudes may be affected by the references to an innovation's developmental phase. Important innovations are covered for many years and during different research stages. For the readers, it may remain unclear as to how long it would take for the innovation to be available in clinical practice. An example is the artificial pancreas, a concept for the treatment of diabetes that has been reported since 1972 [13]. Third, reports on future cure-focused innovations, such as pancreatic cell transplantation for diabetes, may potentially have a stronger impact on the readers' sentiments than news about noncure-focused treatments.

Objectives

To increase the understanding of the associations between the characteristics of news about future treatments for chronic illnesses and the readers' sentiments, we assessed the posts about diabetes research on Facebook pages, together with the corresponding user comments.

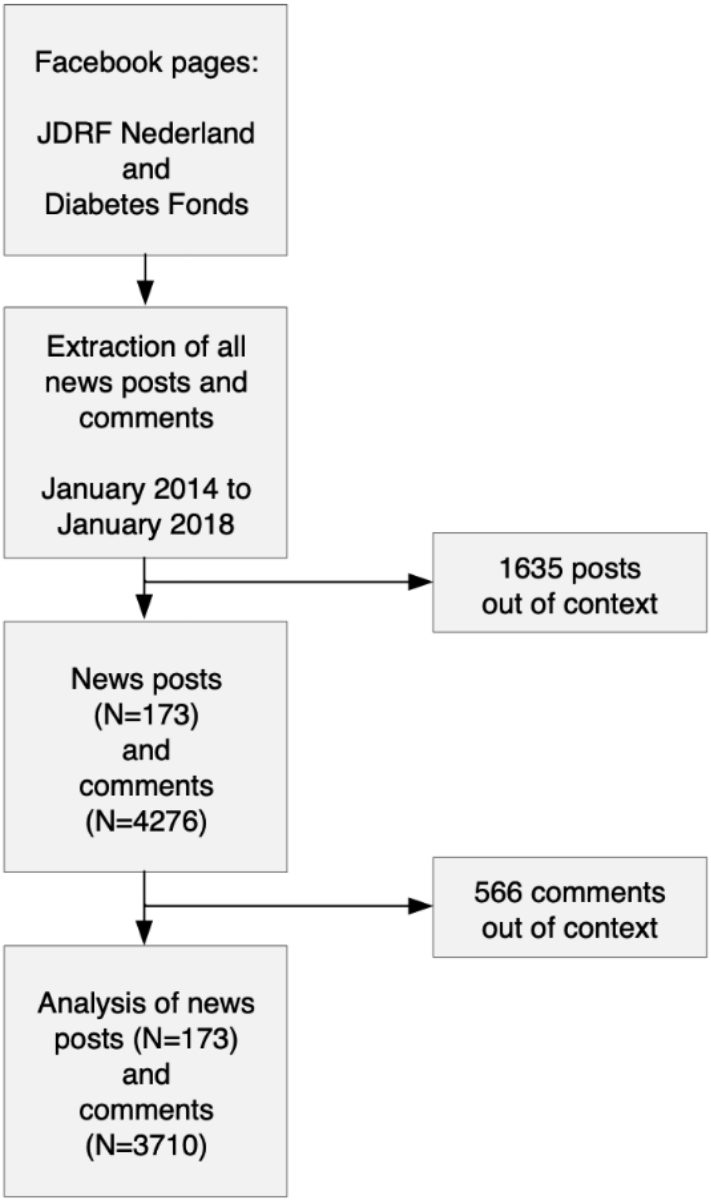
Methods

Data Source

A retrospective observational study was performed on a corpus of diabetes news messages posted on publicly accessible *Facebook pages* between January 1, 2014, and January 1, 2018, and the associated reader comments. Facebook pages enable public figures and businesses to create a public presence on Facebook. Every person on Facebook can connect with these pages by *liking* them, after which they receive updates in their news feed and can interact with them [14]. The 2 most-followed publicly accessible diabetes pages in the Netherlands were selected: (1) Juvenile Diabetes Research Foundation (JDRF) Nederland [15], the Dutch division of an international type 1 diabetes research foundation, with over 27,000 Facebook followers and (2) Diabetes Fonds [16], a Dutch charity funding of research on all types of diabetes, with over 36,000 Facebook followers.

Data extraction and preparation comprised multiple steps (Figure 1). First, all news posts and associated readers' comments were extracted for the 4 years from January 1, 2014, to January 1, 2018. The Facepager tool, version 3.8.2., developed by Jünger and Keyling [17], was used to scrape the publicly available data from the Facebook pages, including all reader comments. Replies were excluded (ie, comments on comments) as their content and sentiments are influenced by the initial comments on the news posts. Second, all nonscience news–related posts were identified and removed from the corpus. Furthermore, 3 criteria for diabetes research posts were applied: (1) it must contain information about the development of an innovative therapy, technique, product, instrument, or insights into preventive behaviors; (2) it must contain a reference to a traceable scientist or scientific institution (including medical companies); and (3) it must refer to an innovation which is not (or not yet) applied in the Dutch standard diabetes care. Therefore, nonmedical and nonscientific topics (eg, personal experiences, practical tips, travel stories, and fundraising) were excluded. The consensus between 2 raters in a subsample of 14.99% (271/1808) of the comments on initial Facebook posts was used to resolve disagreements. The third step was to extract the source message (eg, Web-based newspaper item) whenever a hyperlink was available and to merge it with the post content.

Figure 1. Overview of the data extraction and preparation. JDRF: Juvenile Diabetes Research Foundation.



Data Classification

All user comments were evaluated by 2 raters as to whether they contained sentiments (positive, negative, or both). Focus was put on both textual expressions (ie, words and sentences) and the use of emoticons [18,19]. The initial interrater reliability in a subsample of 15% of the sentiments was high (kappa=0.80). For the remainder of the sample, any disagreements were resolved by consensus and, where uncertainties remained, by a third reviewer.

Dependent Variables: Positive Sentiments and Written Expression of Hopefulness

In the literature, sentiment analyses generally focus on a combined outcome: *sentiment polarity* (ie, coding a single sentiment expression as either positive, negative, or neutral) [20]. As the readers' comments may include a combination of positivity (eg, enthusiasm and hopefulness) and negativity (eg, frustrations about waiting for a long time), both positive and negative sentiments were detected and coded for this study.

The binary dependent variable *positive sentiment* in a comment was manually detected and coded as *present*, when the following paraphrase described the utterance correctly: The Facebook user had the aim to express a positive emotion, attitude, or affective state in reaction to a corresponding post about innovative diabetes treatments. Signal examples were expressions of interest, curiosity, enthusiasm, attraction, desire, admiration, surprise, amusement, hope, excitement, gratitude, thankfulness, joy, elation, triumph, jubilation, patience, and contentment [21-23].

Individuals on the Facebook pages on diabetes frequently express that they *have hopes* or *are very hopeful*, or type *let's hope so!* Hope is a distinct positive sentiment that is considered essential for chronic patients to cope with their disease [24,25]. Having hope can be defined as perceiving a pathway from a negative situation to a favorable state of affairs [26]. Higher

levels of hope were found to be associated with a lower prevalence of diabetes [10] and lower mortality in elderly people with diabetes [11].

A binary hopefulness-estimate was created to put a focus on hope as a distinct and essential positive sentiment. In all Facebook comments, the presence of the following Dutch and English conjugations and adjectives based on the verb *to hope* (Dutch: *hopen*) was programmatically detected by using Python programming language [27]: *hoop*, *hoopt*, *hopen*, *hoopte*, *hoopten*, *gehoopt*, *gehoopte*, *hopelijk*, *hope*, *hoped*, *hoping*, *hopeful*, and *hopefully*.

Dependent Variables: Negative Sentiments

The binary dependent variable *negative sentiment* in a comment was manually detected and coded as *present*, when the following paraphrase described the utterance correctly: The Facebook commenter had the aim to express a negative emotion, attitude, or affective state in reaction to a corresponding post about innovative diabetes treatments. Signal examples were expressions of indifference, habituation, boredom, aversion, disgust, revulsion, alarm, panic, fear, anxiety, dread, anger, rage, sorrow, grief, frustration, disappointment, discontentment, and restlessness [21-23].

The act of expressing and sharing these Facebook post–related emotions reveals an underlying negative sentiment toward the aspects of the innovation or the news message. Commenters sharing, for example, their boredom or indifference on the Web are unlikely to recommend other patients to try the innovation in the future, nor will they follow the news about the therapy actively; positive and neutral reactions would leave these latter 2 behavior options open.

Independent Variables: Intended Therapeutic Effect

Overall, 2 raters identified the intended therapeutic effects of diabetes innovations through discussion and then specified the 5 major categories. These therapeutic effects yielded by diabetes research ranged from simple, practical solutions to a complete cure (Table 1).

Table 1. A total of 5 categories of intended diabetes research effects with examples.

Intended effect	Examples
Prevention	Research into a viral trigger for type 1 diabetes; effects of hygiene; nanotechnology; and early diagnosis
Practical solution	Hypoglycemia alarm watch; an insulin temperature sensor; hypoglycemia watchdog; glucose monitoring app; and diabetic shoe
Symptom control and relief	Cognitive behavioral therapy; research on nephropathy; and research on cardiomyopathy in type 2 diabetes
Glucose regulation	Artificial pancreas; inhalable insulin; Cone Insulin G1; and an insulin delivery system
Diabetes cure	Beta cell encapsulation; viral gene transfer; the discovery of immature beta cells; transplanting pancreatic cells; and the effects of vitamin D

Independent Variables: Developmental Phase

Furthermore, the developmental phases of innovative diabetes therapies were identified. First, the references to research results in the Facebook posts itself were searched for. When a reference was missing, and a hyperlink was available, the source of the news message was examined. In a previous study, our research group distinguished the different research phases, or the levels of evidence, in health news [9]. Health news articles may contain a reference to positive results from (in the increasing order of reliability) (1) observational, often epidemiological, studies, (2) fundamental research on concepts and theories to improve understanding, (3) preclinical (eg, animal studies) and nonclinical studies to support concrete product development, and increasing the chances for clinical trials in humans to start soon, (4) clinical trials in a small sample of humans (eg, phase II pharmaceutical trials), (5) clinical trials in a large population of humans (eg, phase III pharmaceutical trials), and (6) reports on near-market entry whenever a marketing registration has been or soon will be provided by domestic or overseas authorities. The developmental phase was labeled as *not described*, when the research phase was not recognizable either in the Facebook posts or the source message. In total, 2 raters independently scored a subset of 15% of the news posts (kappa=0.92). Consensus was used to resolve disagreement and indistinctness, and 1 rater subsequently coded the remaining 85%.

Independent Variables: Language Intensity

Language intensifiers were defined as words that are used to enhance and give emotional context to the other words that they modify. Literature also refers to such words as the *pars pro toto* superlatives [28,29].

By using Python programming code [27], all words in all Facebook posts were automatically counted and listed in the order of word usage frequency. First, after selecting 2 commonly used designations of the US Food and Drug Administration, *breakthrough* and *promising* [3] (Dutch: doorbraak and veelbelovend), 2 raters discussed and selected the following 15 most frequently used language intensifiers used in the diabetes research news posts: fabulous, beautiful, great, special, important, at last, lifesaving, discovery, dream, positive, powerful, truly, enormously, super, and happy (Dutch: geweldig, mooi, fijn, bijzonder, belangrijk, eindelijk, levensreddend, ontdekking, droom, positief, krachtig, werkelijk, ontzettend, super, and blij). Second, the 17 intensifiers were searched for and counted per Facebook post, and the number was converted into a 3-category variable: *no text intensifiers*, *1 or 2 intensifiers*, and *3 to 9 intensifiers*.

Analysis

The IBM SPSS Statistics program, version 25, was used to evaluate the differences in the probabilities that sentiments (positive, negative, and hopefulness) were reflected in the user

comments, depending on the developmental phase, intended therapeutic effect, and the presence of language intensifiers in the text. Crude and mutually adjusted binary logistic regression models were used to calculate odds ratios (ORs) and 95% CIs. Furthermore, it was assessed whether Facebook pages ID, commenter ID, and gender data contributed to the logistic regression models.

Results

Innovative Methods

Table 2 shows that between January 1, 2014, and January 1, 2018, a total of 173 diabetes news messages about innovative methods to treat diabetes were posted on the 2 largest publicly accessible Facebook pages in the Netherlands. These posts evoked 3710 reader comments, containing a total of 2727 positive, 880 negative, and 363 neutral sentiments and 513 verbal expressions of having hope.

Table 2. The number of extracted news posts and user comments by the innovations' intended therapeutic effect and developmental phase, and the number of comment sentiments.

News characteristics and comment sentiments	News posts, n	User comments, n (%)	Comments per post, mean
Total	173	3710 (100)	21
Intended therapeutic effect			
Disease prevention	16	211 (5.69)	13
Practical solution	14	250 (6.74)	18
Symptom relief	11	182 (4.91)	17
Improved glucose regulation	56	1341 (36.15)	24
Cure	76	1726 (46.52)	23
Developmental phase			
Evidence from fundamental research	47	859 (23.15)	18
Evidence from pretrial phases	32	758 (20.43)	24
Evidence from small human trials	33	712 (19.19)	22
Evidence from large human trials	10	434 (11.70)	43
Near-market entry	14	259 (6.98)	19
Evidence from observational studies	12	185 (4.99)	15
Not described	25	503 (13.56)	20
Comment sentiments			
Positive only	—[a]	2467 (66.50)	—
Negative only	—	620 (16.71)	—
Mixed positive and negative	—	260 (7.01)	—
Positive, including mixed	—	2727 (73.50)	—
Negative, including mixed	—	880 (23.72)	—
Neutral	—	363 (9.78)	—
Verbal expressions of *having hope*	—	513 (13.83)	—

[a]Not applicable.

Sentiments and the Innovation's Intended Therapeutic Effect

First, it was tested whether the news messages about the innovative ways to cure diabetes were associated with different sentiments than the news messages related to other therapeutic effects. Table 3 shows that diabetes prevention (OR 0.55, 95% CI 0.37-0.84), improved blood glucose regulation (OR 0.68, 95% CI 0.56-0.84), and symptom relief (OR 0.31, 95% CI 0.21-0.44) were associated with less positive sentiments as compared with potential diabetes cures. Moreover, Table 3 shows that the analyses of negative sentiments show a similar pattern, although this was only significant in blood glucose regulation (OR 1.38, 95% CI 1.12-1.70; for being associated with more negative sentiments). Table 4 shows that, in line with the readers' positive sentiments, hopefulness was most frequently expressed when Facebook news reported on cure-focused therapies.

Table 3. Logistic regression analysis of the association among 3 characteristics of diabetes news in Facebook posts (mutually adjusted) and positive and negative sentiments in the user comments on Facebook pages.

News characteristics	Facebook posts, n	Positive sentiment			Negative sentiment		
		Yes, n (%)	No, n (%)	OR[a] (95% CI)	Yes, n (%)	No, n (%)	OR (95% CI)
Intended therapeutic effect							
Diabetes prevention	16	94 (3.45)	117 (11.90)	0.55[b] (0.37-0.84)	85 (9.66)	126 (4)	1.48 (0.98-2.25)
Practical solutions	14	196 (7.19)	54 (5.49)	1.01 (0.71-1.44)	52 (5.91)	198 (7)	1.01 (0.71-1.45)
Symptom relief	11	86 (3.15)	96 (9.77)	0.31[b] (0.21-0.44)	49 (5.57)	133 (5)	1.24 (0.84-1.82)
Blood glucose regulation	56	970 (35.57)	371 (37.74)	0.68[b] (0.56-0.84)	343 (38.98)	998 (35)	1.38[b] (1.12-1.70)
Diabetes cure	76	1381 (50.64)	345 (35.10)	1.0[c]	351 (39.89)	1375 (49)	1.0[c]
Developmental phase							
Observational evidence	12	61 (2.24)	124 (12.61)	0.31[b] (0.15-0.66)	79 (8.98)	106 (3.75)	1.88[b] (1.17-3.02)
Fundamental evidence	47	610 (22.37)	249 (25.33)	0.71[b] (0.53-0.95)	238 (27.05)	621 (21.94)	1.22 (0.92-1.63)
Preclinical evidence	32	639 (23.43)	119 (12.11)	1.46[b] (1.07-1.99)	112 (12.73)	646 (22.83)	0.55[b] (0.41-0.76)
Small trial evidence	33	541 (19.84)	171 (17.40)	1.06 (0.80-1.42)	147 (16.70)	565 (19.96)	0.75[b] (0.56-1.00)
Large trial evidence	10	317 (11.62)	117 (11.90)	1.0[c]	115 (13.07)	319 (11.27)	1.0[c]
Near-market entry	14	175 (6.42)	84 (8.55)	0.73 (0.52-1.04)	70 (7.95)	189 (6.68)	1.39 (0.90-2.14)
Not mentioned	25	384 (14.08)	119 (12.11)	1.03 (0.76-1.40)	119 (13.52)	384 (13.57)	0.96 (0.71-1.31)
Language intensifiers[d]							
3-9 intensifiers	21	380 (13.93)	156 (15.87)	0.97 (0.75-1.25)	132 (15.00)	404 (14.28)	1.12 (0.86-1.45)
1-2 intensifiers	55	1120 (41.07)	314 (31.94)	1.13 (0.94-1.35)	334 (37.95)	1100 (38.87)	1.18 (0.99-1.42)
0 intensifiers	97	1227 (44.99)	513 (52.19)	1.0[c]	414 (47.05)	1326 (46.86)	1.0[c]

[a]OR: odds ratio.

[b]Statistically significant odds ratio.

[c]For reference category, CI is not applicable.

[d]Text intensifiers were a nonsignificant addition to this model but were left in the model to answer study questions.

Table 4. Logistic regression analysis of the association among 3 characteristics of diabetes news in Facebook posts (mutually adjusted) and expressed hopefulness (eg, hopefully and I hope) in the user comments on Facebook pages.

News characteristics	News posts, n	Textual expression of hopefulness		
		Yes, n (%)	No, n (%)	OR[a] (95% CI)
Potential therapeutic effect				
Prevention	16	15 (2.92)	196 (6.13)	0.92 (0.50-1.72)
Practical solution	14	10 (1.95)	240 (7.51)	0.18[b] (0.09-0.36)
Symptom relief	11	4 (0.78)	178 (5.57)	0.13[b] (0.05-0.36)
Glucose regulation	56	136 (26.51)	1205 (37.69)	0.49[b] (0.38-0.64)
Diabetes cure	76	348 (67.84)	1378 (43.10)	1.0[c]
Developmental phase				
Observational evidence	12	2 (0.39)	183 (5.72)	0.07[b] (0.02-0.33)
Fundamental evidence	47	133 (25.93)	726 (22.71)	0.97 (0.65-1.43)
Preclinical evidence	32	164 (31.97)	594 (18.58)	1.47[b] (1.01-2.14)
Small trial evidence	33	96 (18.71)	616 (19.27)	1.26 (0.85-1.87)
Large trial evidence	10	46 (8.97)	388(12.14)	1.0[c]
Near-market entry	14	18 (3.51)	241 (7.54)	0.89 (0.49-1.59)
Not described	25	54 (10.53)	449 (14.04)	0.93 (0.61-1.43)
Language intensifiers[d]				
3-10 intensifiers	21	80 (15.59)	456 (14.26)	1.15 (0.84-1.58)
1-2 intensifiers	55	242 (47.17)	1192 (37.28)	1.09 (0.87-1.36)
0 intensifiers	97	191 (37.23)	1549 (48.45)	1.0[c]

[a]OR: odds ratio.

[b]Statistically significant odds ratio.

[c]For reference category, CI is not applicable.

[d]Text intensifiers were a nonsignificant addition to this model but were left in the model to answer the study question.

Sentiment and the Innovation's Developmental Phase

Furthermore, it was examined whether the commenters' sentiments were related to the covered innovations' developmental phases. Tables 3 and 4 show that, compared with the success in the larger patient trials, evidence from the preclinical phases led to more positive sentiments (OR 1.46, 95% CI 1.07-1.99). Earlier observational (OR 0.31, 95% CI 0.15-0.66) and fundamental findings (OR 0.71, 95% CI 0.53-0.95), however, led to less positive sentiments. Tables 3 and 4 also show that this sentiment pattern was, for the most part, confirmed in the analysis of negative sentiments and the expressions of hopefulness.

Sentiment and News Message Language Intensification

It was examined whether the intensification of language was associated with sentiments and hopefulness. However, Tables 3 and 4 show that there were no significant relationships between the language intensification of Facebook posts about diabetes research and the sentiments (positive, negative, and hopefulness) of those who reacted to it on Facebook.

Controlling for Other Variables

To verify the robustness of our findings, additional variables and levels were tested for any effect on our regression analysis. It was found that the Facebook pages ID (JDRF Nederland vs Diabetes Fonds) did not contribute to the regression models. Furthermore, data on the commenters' gender were available for 71% (2634/3710) of the comments with identifiable sentiment (the first batch of 2 data extractions). Analysis of this sample showed that controlling for gender did not greatly alter the patterns and magnitudes of our results. In addition, the necessity to include commenter ID as a level in our regression model was rejected owing to the flat distribution of comments by the commenters in the same subsample: 80% (1495/1870) of the commenters commented once only.

Our final model only contained the 3 main independent variables: therapeutic effect, developmental phase, and language intensification.

Discussion

Principal Findings

In this analysis of 4 years of Facebook posts and comments on diabetes news and user sentiments, posts about potentially curative innovations were associated with more positive general sentiments than the posts not about potential cures, as expected.

However, unexpectedly, innovations supported by evidence from phases just before human clinical trials showed the strongest positive association with improved general sentiments. The observational research results were associated with the most negative general sentiments in the user comments. In addition, and contrary to our expectations, this study found no evidence for the associations between language intensity and the readers' sentiments.

Explanation of Findings

A strong positive association was found between cure-focused innovations and positive sentiments. The explanation for this finding is likely to be the absence of cure-focused therapies, to date, for the debilitating disease that diabetes still is. However, negative sentiments may also be provoked by cure-focused innovation, for example, when frustrations about perceived false promises have the upper hand. The strong negative association between the written expression of hopefulness and the news about noncure-focused innovations suggests that the concept of hopefulness only plays a role regarding the desire for a cure.

When looking at the developmental phases, general sentiments were most positively associated with positive results in the preclinical phases closely before human trials. This finding conflicts with the scientific standard that the proof of concept is demonstrated by doing randomized clinical trials. Although it was not assessed by us, an explanation for the preclinical positivity may be the overly optimistic way in which news outlets frequently cover animal studies [30].

Negative associations of the fundamental research with positive sentiments can be explained using construal level theory [31]. First, as the success of therapies in the earliest research stages is far away in time, the patient's thinking about these innovations becomes more abstract and the consequent anticipation may decrease. Frequently, the medical applicability of very early stage therapies is indeed abstract. A second explanation may be that bad personal experiences, with waiting for other cure-focused innovations, affect the so-called experiential distance (ie, perception of the chance that treatment may become a reality, based on earlier experiences).

Specific research topics may explain the strong association between observational research and the less positive general sentiments. Both disbelief and powerlessness in readers may have arisen from Facebook posts that describe how patients *should have behaved in the past* to prevent their disease. Furthermore, the association between the less positive general sentiments and the news about near-market innovations may be related to fears and frustrations regarding low availability and the doubt on medical insurance coverage.

A possible explanation for the absence of associations between language intensifiers in the news content and the sentiments may have its origin in the characteristics of our target population. Patients and others interested in diabetes seemed to be able to distinguish between the objective content and the subjective use of language. People commenting on the investigated diabetes pages are involved in the burdens of the chronic disease—enduring much and awaiting therapies for many years—and their emotions may not (or may no longer) be as affected by the subjective language.

Implications of This Study

Previous studies suggest that exaggeration of medical research is a problem [1-7]. Our study shows no evidence that language intensifiers are associated with the sentiments of the online diabetes populations. However, improved positive sentiments were found regarding the preclinical trials—just before evidence from humans—that give reason to suspect an undesired effect of health news exaggeration. The Facebook comments were enthusiastic and full of hope, despite the fact that about 88% of the pharmaceutical developments that reach the first human trials will never reach the phase of market approval [8].

These findings suggest that exaggeration is not limited to language intensification and other verbal inflation of research findings. It is the sheer coverage frequency of specific health research that may lead to positivity and hope, which is not always justified. The mental shortcut *availability heuristic* relies on the immediate examples that come to a person's mind, possibly putting too much weight on medical information when they read about it frequently [32]. At the same time, the importance of having hope when suffering from a chronic disease must not be underestimated. Hope mediates the relationship between psychological distress and health status and is an essential factor to cope with a disease [24,25].

High-quality health information is increasingly important. Responsible news authors must give context, interpret scientific findings, filter what is important to their target group, and act as an honest and valid conduit, especially as the role of social media is increasing every year [33]. Furthermore, the specific importance of social media to patients must be emphasized. Platforms such as Facebook or Twitter provide tailored information, increase the accessibility of news, and function as social and emotional peer supporters [34-36]. Being well-informed about scientific developments fulfills an essential need for the health information–monitoring patient [37], and it is known that the patients' subjective well-being also clearly and positively affects health and all-cause mortality [38].

Strengths and Limitations in Comparison With Other Studies

To our knowledge, our study is the first to quantitatively investigate the associations between the health news characteristics and the sentiments of readers dealing with chronic illnesses.

Moreover, an extra focus was put on the written expressions of being hopeful, enabling the confirmation of general sentiment associations in 1 specific disease-related sentiment. By assessing common language intensifiers, it was possible to differentiate

between the objective characteristics of the news posts and a subjective language component. Our large sample size enabled us to mutually adjust the 3 news-related variables. Moreover, the validity of our sentiment outcome increased as offhand comments were observed, written down in an unforced situation. By using a Python regular expression search, the reliability of finding all language intensifiers was high. In addition, the kappa values for rating sentiments and coding message characteristics were high, and coding consensus was achieved regarding occasional discrepancies.

This study does, however, have limitations. First, it is not known what our population's exact proportion sizes are regarding the patients, their social context, and others who were perhaps only momentarily interested in diabetes and left a comment. Moreover, although the language intensifiers were included, other journalistic language elements, such as emotionalization of news, were not included as a potentially predicting or modifying factor. One final limitation of the study may be the bias that theoretically occurs when readers with either very strong (rejecting) or neutral sentiments refrain from commenting owing to the sentiment itself.

Conclusions

By observing the news posts and comments on diabetes research on 2 large Dutch Facebook pages, we found that the readers' sentiments are associated with both the innovations' developmental phase and the intended therapeutic effect. However, no evidence was found on the association between sentiments and the presence of commonly used language intensifiers in the health news texts.

Our finding that comments on diabetes news on Facebook have the most positive sentiment, and most frequently express hopefulness when clinical efficacy is not yet proven, suggests that the news authors and editors must exercise caution when acting as a conduit for medical research news. More experimental research is necessary, in various populations, to determine a healthy balance between being optimally informed and avoiding having false hope.

Conflicts of Interest

None declared.

References

1. Sumner P, Vivian-Griffiths S, Boivin J, Williams A, Venetis CA, Davies A, et al. The association between exaggeration in health related science news and academic press releases: retrospective observational study. Br Med J 2014 Dec 9;349:g7015 [FREE Full text] [doi: 10.1136/bmj.g7015] [Medline: 25498121]
2. Coyne JC, Tennen H. Positive psychology in cancer care: bad science, exaggerated claims, and unproven medicine. Ann Behav Med 2010 Feb;39(1):16-26 [FREE Full text] [doi: 10.1007/s12160-009-9154-z] [Medline: 20146038]
3. Krishnamurti T, Woloshin S, Schwartz LM, Fischhoff B. A randomized trial testing us food and drug administration 'breakthrough' language. JAMA Intern Med 2015 Nov;175(11):1856-1858. [doi: 10.1001/jamainternmed.2015.5355] [Medline: 26390245]
4. Dudo A. Scientists, the media, and the Public Communication of Science. Sociol Compass 2015;9(9):761-775. [doi: 10.1111/soc4.12298]
5. Ioannidis JP, Stuart ME, Brownlee S, Strite SA. How to survive the medical misinformation mess. Eur J Clin Invest 2017 Nov;47(11):795-802. [doi: 10.1111/eci.12834] [Medline: 28881000]
6. Schwitzer G. How do US journalists cover treatments, tests, products, and procedures? An evaluation of 500 stories. PLoS Med 2008 May 27;5(5):e95 [FREE Full text] [doi: 10.1371/journal.pmed.0050095] [Medline: 18507496]
7. Goldacre B. Preventing bad reporting on health research. Br Med J 2014 Dec 9;349:g7465. [doi: 10.1136/bmj.g7465] [Medline: 25498123]
8. Hay M, Thomas DW, Craighead JL, Economides C, Rosenthal J. Clinical development success rates for investigational drugs. Nat Biotechnol 2014 Jan;32(1):40-51. [doi: 10.1038/nbt.2786] [Medline: 24406927]
9. Vehof H, Sanders J, van Dooren A, Heerdink E, Das E. Clinical evidence vs preliminary speculation in newspaper coverage of diabetes innovations: a quantitative analysis. Public Health 2018 Jul;160:49-51. [doi: 10.1016/j.puhe.2018.03.022] [Medline: 29734012]
10. Richman LS, Kubzansky L, Maselko J, Kawachi I, Choo P, Bauer M. Positive emotion and health: going beyond the negative. Health Psychol 2005 Jul;24(4):422-429. [doi: 10.1037/0278-6133.24.4.422] [Medline: 16045378]
11. Moskowitz JT, Epel ES, Acree M. Positive affect uniquely predicts lower risk of mortality in people with diabetes. Health Psychol 2008 Jan;27(1S):S73-S82. [doi: 10.1037/0278-6133.27.1.S73] [Medline: 18248108]
12. Abola MV, Prasad V. The use of superlatives in cancer research. JAMA Oncol 2016 Jan;2(1):139-141. [doi: 10.1001/jamaoncol.2015.3931] [Medline: 26512913]
13. Bessman SP, Schultz RD. Sugar electrode sensor for the 'artificial pancreas'. Horm Metab Res 1972 Nov;4(6):413-417. [doi: 10.1055/s-0028-1094020] [Medline: 4655496]
14. Facebook. About - Pages URL: https://www.facebook.com/help/282489752085908?helpref=hc_global_nav [accessed 2019-02-28]
15. Juvenile Diabetes Research Foundation. 2018. URL: https://www.jdrf.nl/ [accessed 2019-04-24] [WebCite Cache ID 77s8TmRV3]

16. Diabetes Fonds. 2018. URL: https://www.diabetesfonds.nl/home [accessed 2019-04-24] [WebCite Cache ID 77s8qpwff]
17. Jünger J, Keyling T. WebCite. Facepager URL: http://www.webcitation.org/77s6MGrQD [accessed 2019-04-24] [WebCite Cache ID 77s6MGrQD]
18. Danesi M. The Semiotics of Emoji: The Rise of Visual Language in the Age of the Internet. New York, USA: Bloomsbury Publishing; 2016.
19. Miller HJ, Thebault-Spieker J, Chang S, Johnson I, Terveen L, Hecht B. 'Blissfully happy' or 'ready to fight': Varying Interpretations of Emoji. In: Proceedings of the 10th International AAAI Conference on Web and Social Media. 2016 Presented at: ICWSM'16; May 17-20, 2016; Cologne, Germany.
20. Saif H, He Y, Alani H. Semantic Sentiment Analysis of Twitter. In: Proceedings of the 11th international conference on The Semantic Web - Volume Part I.: November 11-15, 2012; 2012 Presented at: ISWC'12; 2012 Nov 11; Boston, MA p. 508-524.
21. Robinson DL. Brain function, emotional experience and personality. Neth J Psychol 2008;64(4):152-168. [doi: 10.1007/bf03076418]
22. Schmidt SJ. What does emotion have to do with learning? Everything!. J Food Sci Educ 2017 Jul;16(3):64-66. [doi: 10.1111/1541-4329.12116]
23. Lamm D. Semantic Scholar. Patterns of pleasure--the design and development of customer satisfaction URL: https://www.semanticscholar.org/paper/Patterns-of-pleasure--the-design-and-development-of-Lamm/7deaa62cf465a5b3967834a98be1f294963adb23 [accessed 2019-09-26]
24. Dorsett P. The importance of hope in coping with severe acquired disability. Aust Soc Work 2010;63(1):83-102. [doi: 10.1080/03124070903464293]
25. Rustøen T, Cooper BA, Miaskowski C. The importance of hope as a mediator of psychological distress and life satisfaction in a community sample of cancer patients. Cancer Nurs 2010;33(4):258-267. [doi: 10.1097/NCC.0b013e3181d6fb61] [Medline: 20467303]
26. Snyder C, Forsyth D. Handbook Of Social And Clinical Psychology: The Health Perspective. Oxford, UK: Pergamon Press; 1991.
27. van Rossum G, Drake F. An Introduction to Python. Bristol: Network Theory Ltd; 1995.
28. Burgers C, de Graaf A. Language intensity as a sensationalistic news feature: The influence of style on sensationalism perceptions and effects. Communications 2013;38(2):167-188. [doi: 10.1515/commun-2013-0010]
29. Bowers JW. Some correlates of language intensity. Quart J Speech 1964 Dec;50(4):415-420. [doi: 10.1080/00335636409382688]
30. Adams JU. Health News Review. 2018. Animal Studies, Reported Well, Include Careful Framing and Clear Caveats URL: https://www.healthnewsreview.org/2018/01/animal-studies-reported-well-include-careful-framing-and-clear-caveats/ [accessed 2019-04-24] [WebCite Cache ID 77s9nXqMU]
31. Trope Y, Liberman N. Construal-level theory of psychological distance. Psychol Rev 2010 Apr;117(2):440-463 [FREE Full text] [doi: 10.1037/a0018963] [Medline: 20438233]
32. Berry D. Risk, Communication and Health Psychology. UK: McGraw-Hill Education; 2004.
33. Pew Research Center. 2018. Social Media Fact Sheet URL: https://www.pewinternet.org/fact-sheet/social-media/ [accessed 2019-06-15]
34. Moorhead SA, Hazlett DE, Harrison L, Carroll JK, Irwin A, Hoving C. A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. J Med Internet Res 2013 Apr 23;15(4):e85 [FREE Full text] [doi: 10.2196/jmir.1933] [Medline: 23615206]
35. Petrovski G, Zivkovic M, Stratrova SS. Social media and diabetes: can Facebook and Skype improve glucose control in patients with type 1 diabetes on pump therapy? One-year experience. Diabetes Care 2015 Apr;38(4):e51-e52. [doi: 10.2337/dc14-2487] [Medline: 25805869]
36. Zhang Y, He D, Sang Y. Facebook as a platform for health information and communication: a case study of a diabetes group. J Med Syst 2013 Jun;37(3):9942. [doi: 10.1007/s10916-013-9942-7] [Medline: 23588823]
37. Miller SM. Monitoring versus blunting styles of coping with cancer influence the information patients want and need about their disease. Implications for cancer screening and management. Cancer 1995 Jul 15;76(2):167-177. [doi: 10.1002/1097-0142(19950715)76:2<167::aid-cncr2820760203>3.0.co;2-k] [Medline: 8625088]
38. Diener E, Chan M. Happy people live longer: subjective well-being contributes to health and longevity. Appl Psycol 2011;3(1):1-43. [doi: 10.1111/j.1758-0854.2010.01045.x]

Abbreviations

JDRF: Juvenile Diabetes Research Foundation
OR: odds ratio

Edited by G Eysenbach; submitted 01.05.19; peer-reviewed by E Da Silva, S Oser; comments to author 25.05.19; revised version received 20.07.19; accepted 14.08.19; published 13.11.19.

Please cite as:
Vehof H, Heerdink E, Sanders J, Das E
Associations Between Characteristics of Web-Based Diabetes News and Readers' Sentiments: Observational Study in the Netherlands
J Med Internet Res 2019;21(11):e14554
URL: https://www.jmir.org/2019/11/e14554
doi:10.2196/14554
PMID:31719025

Original Paper

Barriers and Enablers Affecting Successful Implementation of the Electronic Health Service Sisom: Multicenter Study of Child Participation in Pediatric Care

Petra Svedberg[1*], PhD, Prof Dr; Susann Arvidsson[1*], PhD; Ingrid Larsson[1*], PhD; Ing-Marie Carlsson[1*], PhD; Jens M Nygren[1*], PhD, Prof Dr

Halmstad University, Halmstad, Sweden

[*]all authors contributed equally

Corresponding Author:
Petra Svedberg, PhD, Prof Dr
Halmstad University
Box 823
Halmstad, S-30118
Sweden
Phone: 46 35167693
Email: petra.svedberg@hh.se

Abstract

Background: Children's participation in health care is one of the most important components in the management of their disease. Electronic health (eHealth) services that are adapted to the needs of children have the potential for restructuring how children and professionals work together. Therefore, a digital interactive assessment and communication tool, Sisom, was developed to give children aged between 6 and 12 years a voice in their own health care. However, the implementation of eHealth services such as Sisom in daily practice in pediatric health care is rarely investigated.

Objective: The aim of this study was to explore the process of implementing Sisom for children in pediatric care in Sweden. More specifically, the study aimed to (1) evaluate whether the implementation strategy was conducted as planned, (2) understand the barriers and facilitators of the implementation strategy in pediatric care settings, (3) gain insight into how professionals work with the specific intervention, and (4) gain insight into the usefulness and effects of the intervention from the professionals' perspectives.

Methods: A process evaluation design was used to study the implementation of Sisom at 4 pediatric care centers in Sweden. An extensive amount of qualitative and quantitative data was collected before, during, and after the intervention through self-report checklists, memos, and interviews with professionals. In total, 46 children, aged between 6 and 13 years, participated. The children used Sisom on two occasions during 6 months. When they used Sisom, a printed report formed the basis for a forthcoming dialogue between professionals, children, and their parents.

Results: To our knowledge, this is the first implementation study of an eHealth communication tool aimed at strengthening children's participation in pediatric health care. Key factors for successful implementation were alignment of the solution with the values and goals of the organization, health care professionals' beliefs in the usefulness and usability of the solution, and health care professionals' willingness to change their professional roles guided by the solution.

Conclusions: The results from the study show that it is possible to restructure health care delivery toward a child-centered approach, if there is a willingness and preparedness in the organization to implement an eHealth solution with the aim of restructuring the way of working with children's participation.

(J Med Internet Res 2019;21(11):e14271) doi:10.2196/14271

KEYWORDS

children; pediatrics; eHealth; health care; quality improvement; diffusion of innovation; implementation science; participatory medicine

Introduction

According to the United Nation's Convention of the Rights of the Child (UNCRC), all children, regardless of gender, age, background, and disability, have the right to be heard in all matters concerning them [1]. When the UNCRC will be incorporated in Swedish law starting in January 2020, this right will apply to society in general, as well as to the child's everyday life and when they need health care [2]. Moreover, the Swedish Patient Act [3] aims to strengthen the patient's position through participation in their own health care. In spite of a general acceptance concerning the importance of participation, challenges remain in translating such ambitions into practice, especially when it comes to children [4]. Furthermore, the development and implementation of initiatives that promote children's participation in health care is an underdeveloped area [5,6].

Background

Children's participation in health care is one of the most important components in the management of their disease and can have positive effects on the communication among children, parents, and professionals and for participation in decision processes [7]. However, their ability to participate in their own care needs to be strengthened [8-13]. It is largely limited by their restricted ability to convey their needs, expectations, and values and that communication channels focus on the health care provider and the parent [14]. This can result in increased fears and anxieties, reduced self-esteem, depersonalization, and feeling of being unprepared for procedures [15]. It is therefore important to improve the involvement of children in their own health care at a level commensurate with their experience, age, and abilities. Digital communication tools that are adapted to the needs of children have the potential to restructure how children and professionals work together by facilitating children's opportunities and capabilities to describe their experiences and preferences and supporting their role in goal formulation and decision making [13,16]. Despite the potential of such tools, research on their implementation in clinical practice has been given less attention [5,8,17].

The use of electronic health (eHealth) services for health issues in general in pediatric health care is scarce, and few eHealth services focus on enabling and supporting children and young people's participation in pediatric health care. Most of the eHealth services that have been developed are related to children with mental illnesses [18-20], blood disorder [21], or cancer [22] and are primarily focusing on symptom assessment, medication adherence, information, training, and self-management. We have only found one review of eHealth services that was designed to support the communication between children and health professionals with the overall purpose to strengthen children's and young people's participation in health care [22]. Furthermore, the evidence base to guide policy and practice for eHealth services targeting children's participation in health issues is insufficient [19-21,23-26].

Sisom is an eHealth service [27] that was developed to give children aged between 6 and 12 years *a voice* to support their involvement in their own health care [16,28,29]. Previous research conducted on Sisom has mainly focused on issues in the design and pretesting phases around the development and evaluation of content validity, usability, and concurrent validity [15,29-35]. However, the implementation of Sisom in daily practice in pediatric health care has never been investigated. To narrow the gap between research and application in clinical practice, a thorough understanding of the implementation processes of Sisom in clinical practice is needed. This project deals with the implementation of Sisom for children in pediatric health care with the overall goal of strengthening their participation in their own health care. The implementation of research results in clinical practice is, however, in general, a challenging process, [36] and the evidence base for eHealth services is modest [37]. Therefore, the results from this study will provide valuable knowledge and guidance for policy and practice for general implementation of eHealth services in pediatric health care.

Aim

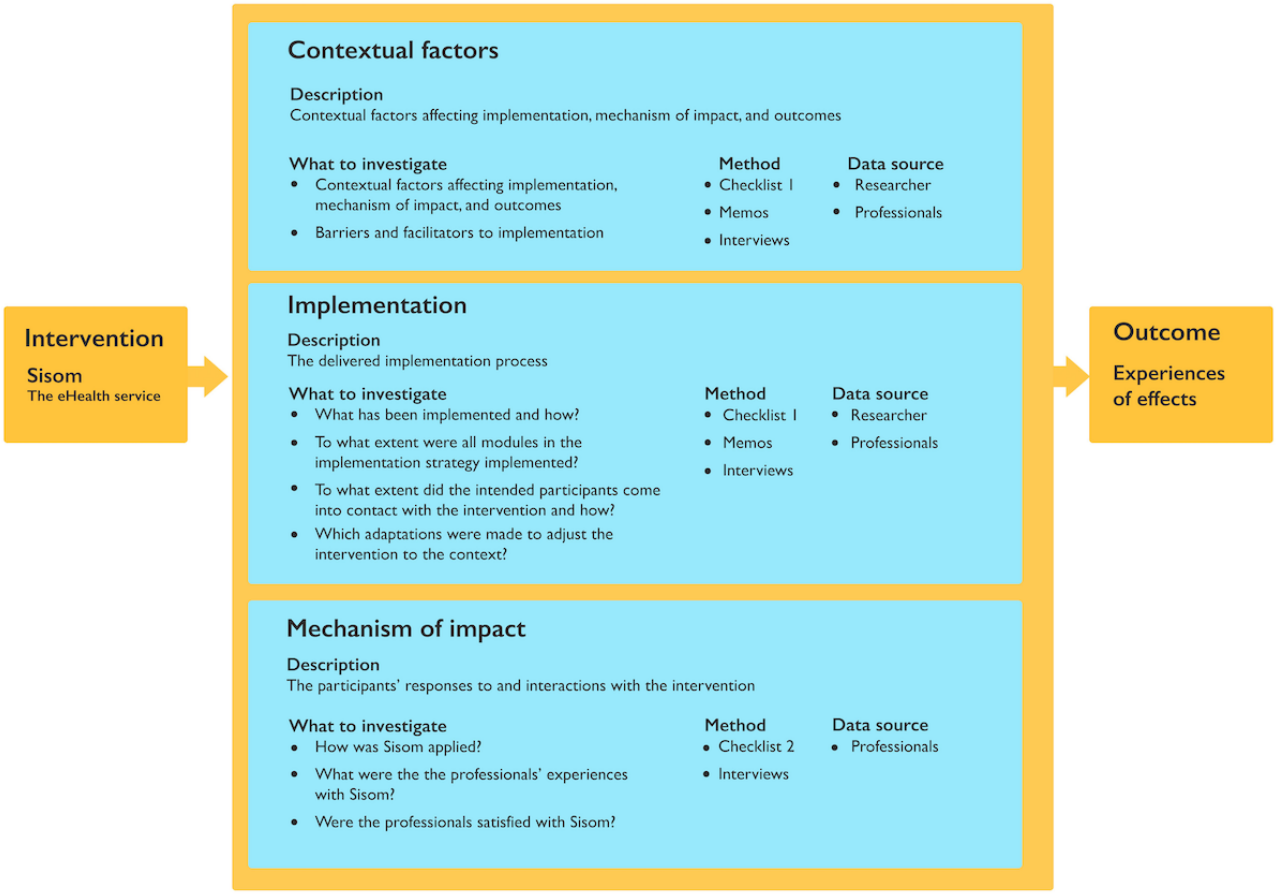
The aim of this study was to explore the implementation process of the eHealth service Sisom for children in pediatric care in Sweden. More specifically, the process evaluation attempts to (1) evaluate whether the implementation strategy was conducted as planned, (2) understand the barriers and facilitators of the implementation strategy in pediatric care settings, (3) gain insight into how professionals work with the specific intervention, and (4) gain insight into the usefulness and effects of the intervention from the professionals' perspectives.

Methods

Design

The implementation of an eHealth service for children in health care was studied through a process evaluation design (Figure 1) using both quantitative and qualitative methods [38]. All procedures in the study were performed with permission from the Regional Ethical Review Board in the south of Sweden (Reg. No. 2015/31).

Figure 1. Model describing the process evaluation blueprint. eHealth: electronic health.



Study Setting and Participants

A total of 4 pediatric care centers at 3 different hospitals in Sweden were included in the study during 2016-2017. The pediatric care centers differed somewhat in terms of hospital size, care delivery processes, and the range of diagnoses treated at the center (Table 1).

Table 1. Study setting and participants.

Centers	Type of care	Patient group	Participants
Center A (large university hospital)	Counselor at the out- and inpatient care centers working with advice and support to facilitate everyday life for families with children with chronic illness.[a]	Children (aged 0-17 years) mostly treated for hematologic diseases and HIV infections	3 counselors
Center B (large university hospital)	Counselor at the out- and inpatient care centers working with advice and support to facilitate everyday life for families with children with chronic illness.[a]	Children (aged 0-17 years) mostly treated for various forms of cancer, diabetes, and heart diseases	10 counselors
Center C (small rural hospital)	Pediatric oncology outpatient care center with a team of pediatricians, nurses, counselors, physiotherapists, and dieticians	Children (aged 0-17 years) mostly treated for various forms of cancer	2 nurses
Center D (small rural hospital)	Pediatric neurology outpatient care center with a team of pediatricians, nurses, counselors, physiotherapists, and dieticians	Children (aged 0-17 years) treated for neurologic diseases	2 nurses

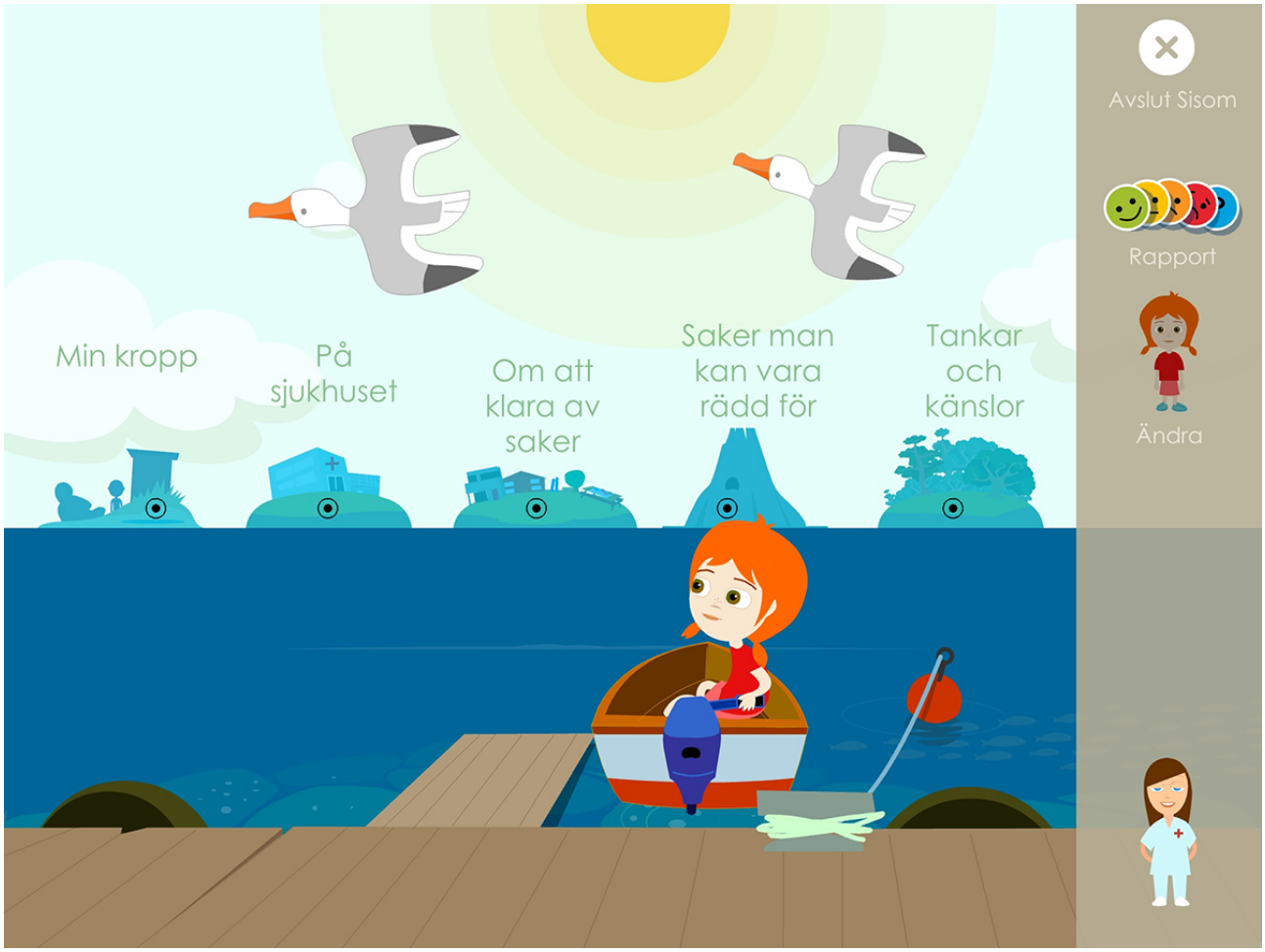
[a]At the outpatient care centers, the children usually only met the counselors, but sometimes additional professionals from the team were involved during the appointment. At the inpatient care centers, the counselors were part of a team of professionals, including pediatricians, nurses, physiotherapists, occupational therapists, psychologists, dieticians, and pedagogues.

The Electronic Health Service, Sisom

The intervention was based on an eHealth service for children in pediatric care called Sisom (Norwegian acronym for *Tell it how it is*). Sisom was developed with and for children aged between 6 and 12 years with the purpose of giving them a *voice* in their own care [16,29]. With its child-friendly interface on mobile devices, Sisom uses spoken language, text, sounds, animations, and intuitively meaningful metaphors and pictures to represent different life situations and symptoms, allowing a wide range of children to understand the meaning and communicate. In the form of a self-designed avatar, the child

goes on a virtual journey traveling from island to island. Sisom currently consists of 5 islands: at the hospital, about managing things, my body, thoughts and feelings, and things one can be afraid of (Figure 2). The islands contain a total of 84 questions related to dimensions relevant to describing the children's life situation and symptoms. The questions are presented in both text and verbally by the system. Thus, if any children had difficulties reading the text, they are guided by the audio recording of the questions and instructions. Children report their self-assessments and feelings in relation to these dimensions by selecting the level of agreement on a 5-point Likert scale with cartoon faces (with different colors and expressions). In addition, the children can specify areas of pain, bruises, and rash on a body map. When the children have gone through all the islands in Sisom, a report is printed out with the answers to the various questions. This report then forms the basis for a forthcoming dialogue between professionals and children (and sometimes parents).

Figure 2. Overview of the electronic health service Sisom.



Implementation Strategy

The implementation strategy used consists of different components to support the implementation and maintenance of the eHealth service in the participating pediatric centers. The strategy was administered in the form of a 12-month program including (1) the introduction and anchoring of the project via physical meetings and email correspondence, (2) a 1-day workshop, (3) intervention period, (4) continuous facilitation, and (5) continuous support from the university, and finally (6) follow-up seminars (Table 2). In this project, the role of the local facilitators was to support and communicate with the professionals working with the intervention at each center to facilitate implementation [39].

Table 2. Content of the implementation strategy.

Implementation strategy components	Content	Time
Introducing and anchoring via physical meetings, email, and phone	Introducing managers and professionals to the project: discussion about how to ideally organize the project; introduction to the concept of child participation; presentation of facilitator's role; preparation of written information to professionals and managers.	2-6 months before the workshop
Half-day or one-day workshop educating professionals and facilitators	Workshop on evidence on child participation in health care; introduction to the concept of child participation; introduction and practicing using Sisom; discussion on contextual barriers and facilitators; introducing the data collection.	Starting point
Intervention	Children and professionals use Sisom during a period of 6 months.	1-6 months
Continuous facilitation	Local facilitators support professionals in the implementation.	1-6 months
Continuous supervision	Supervisors from the university support professionals in the implementation.	1-6 months
Follow-up dialogue	Dialogue with professionals providing opportunities to share experiences of using Sisom and to discuss contextual barriers and facilitators.	1-2 months after the workshop

Data Collection

An extensive amount of qualitative and quantitative data was collected before, during, and after the intervention through self-report checklists and interviews with professionals. The researchers also wrote memos when they were in contact with the centers and used these as data for analysis.

A total of 2 different self-report checklists were used during the project: after the education day (checklist 1) and directly after using Sisom with the child (checklist 2). The research team developed these checklists based on the components in the blueprint (Figure 1) to capture context, fidelity, dose delivered and reach, as well as professionals' compliance with the content of the education and how they worked with the intervention.

Individual interviews or interviews in pairs were performed with 14 of the 17 professionals participating in the intervention and 1 manager at the university hospital. A total of 4 researchers conducted interviews with the participants and only with participants who they had not previously met and from centers where they had not previously participated in the implementation process. A semistructured open-ended interview guide was developed to encourage the participants to speak openly about their experiences (related to the questions in Figure 1). The interviews were audio recorded and transcribed verbatim.

Data Analysis

Transcribed interviews, memos, and free-text responses from checklists 1 and 2 were analyzed using qualitative content analysis [40]. All transcripts were first read several times and deductively coded (by 2 of the authors) into the process evaluation components: contextual factors, implementation, mechanism of impact, and outcome [38]. An inductive coding followed, organizing the preliminary codes in potential subcategories. The analytical process was discussed continuously within the research team to increase the trustworthiness and rigor of the analysis. The data derived from checklist 2 were analyzed using descriptive statistics in SPSS 25.0 (IBM).

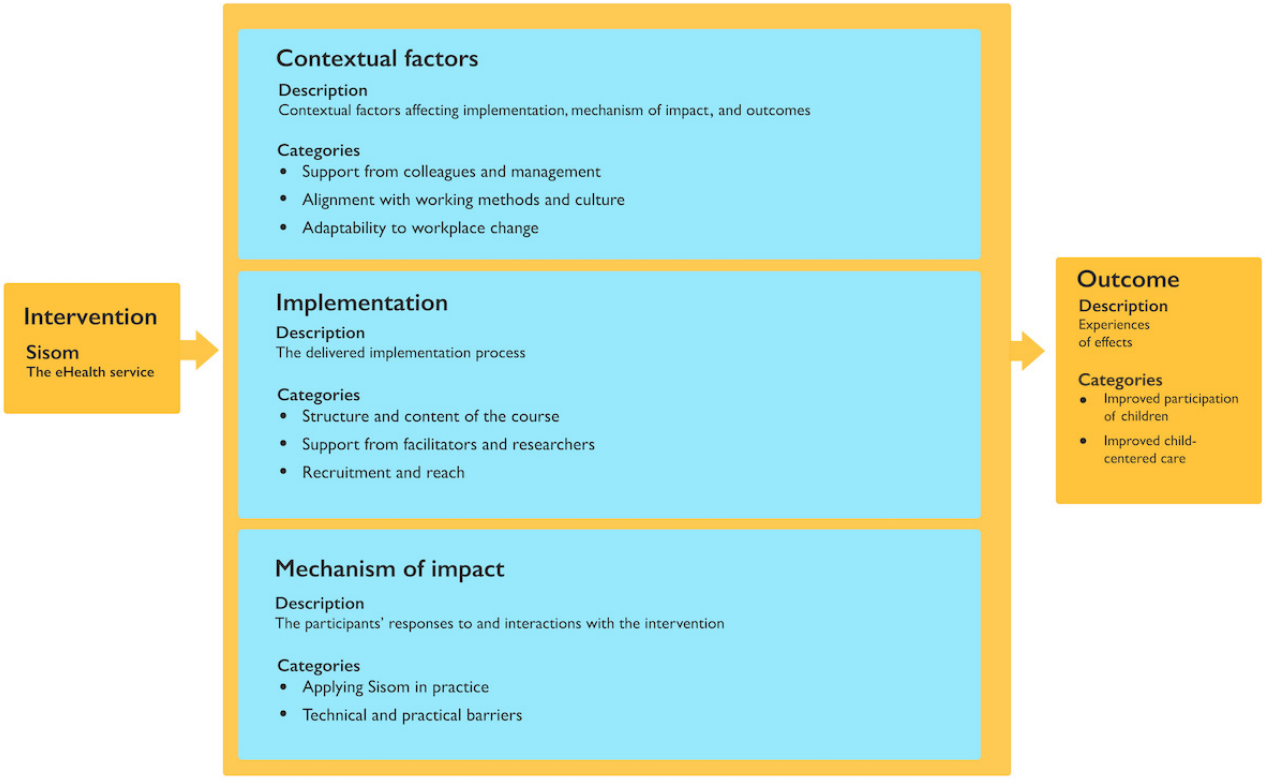
Ethical Considerations

The Regional Ethical Review Board approved the study (Reg. No.: 2015/31), and the study conforms to the principles outlined in the Declaration of Helsinki [41]. All participants received written and verbal information about the study. The participants were given information about the voluntary nature of the study, confidentiality, and that they were free to withdraw their consent at any time during the process. The participants gave written consent to participate and were free to choose the time and place of the interviews.

Results

The process evaluation contained 4 key components according to the evaluation blueprint: contextual factors, implementation, mechanisms of impact, and experiences of effects (Figure 3).

Figure 3. A model describing the categories and subcategories resulting from analysis of data acquired during and after implementation of Sisom according to the process evaluation blueprint. eHealth: electronic health.



Contextual Factors

A total of 3 subcategories emerged from the data: *support from colleagues and management*, *alignment with working methods and culture*, and *adaptability to workplace change*.

Support From Colleagues and Management

A decisive factor for implementing Sisom at the centers was described as the management's permission and support for carrying out the project. All participants experienced that they had support from the management to implement the project, although anchorage varied between the centers. One challenge was that hospitals often have many ongoing projects, which results in competition for resources and prioritization between projects. However, in this project, the management saw the project as part of quality development in relation to strengthening of child participation at the center and thus in line with the vision of the hospital around value-based care and innovation. At one of the hospitals, a written agreement between the researchers and the hospital was important for anchoring and giving the project legitimacy in the organization:

> *Yes, I think so (have had management support). Just saying that it was OK to participate. That it's OK to write a contract about Sisom combining the perspectives of the university and the hospital. They've been open and accessible for...they've seen it as a development (of the services).* [Informant 10]

Another important factor described by the professionals was that their colleagues were interested and had a positive attitude toward implementing Sisom at the center. The professionals experienced that their colleagues were curious about how Sisom can support their work with child-centered care at an organizational level. The colleagues were also interested in knowing what the output was from children's usage of Sisom and anticipated it would provide important information for the care team meetings and medical rounds at the center:

> *...I've not noticed any resistance to this. There hasn't been any. Not among the health care staff or in our own organization...there's been more of a curiosity about what we've been doing.* [Informant 6]

A frustration that only a few professionals at each center were practically involved in the implementation of Sisom was spoken of. Those expressing this wished that Sisom had been available to all professionals allowing its use in all situations at the center to fully support the children in their care. With only a limited number of professionals working with Sisom, the feeling at the center was that some counselors/nurses knew things about the child that other professionals would have liked to know.

Alignment With Working Methods and Culture

The common commitment among professionals to work consistently toward increasing child participation at each center and the shared value base of the professionals toward

child-centered care were a contextual factor that promoted the implementation of Sisom. The professionals stated that it was the parents' perspectives that predominated and had influence over the child's care. There were no suitable tools for conveying the children's perspectives on their care situation, and most of the professionals talked of a desire to have instruments that could help them in their conversations with the children. Mapping and adjusting to traditions, values, and attitudes of the workplaces was important for the implementation process. It was also described as easier to implement Sisom in a center where professionals had more control over their own time and where the child met the same professionals each time they visited the hospital. The professionals described that Sisom has an obvious place in their work:

> *...something that the social worker should do is to talk to the children but then there are several health centers where the social worker's job is to talk to the parents instead. And at a health center where it is more common to do so (talk to the children) then Sisom can be even more effective and useful instrument. Because one gets access to the child...* [Informant 2]

A contextual barrier for the successful implementation of Sisom was a lack of time and that everyday work was in many ways interrupted by emergency events. The professionals were not always able to allocate enough time for meeting with the child because of the planning of additional meetings with other professionals during the visit at the hospital. The professionals described that they did not want to *squeeze* Sisom into their daily work if they were unable to guarantee that they could provide the prerequisites for carrying it out to a high standard for all those involved, the children and their families and the professionals themselves. Some professionals experienced that there was no space for working in a structured manner with Sisom:

> *...it's difficult to implement everything...a great deal is needed for implementing different things in a health service like ours, where things happen all the time, where emergencies can dominate somewhat and you have to put aside your almanac all the time. Then a great deal is needed. A lot of involvement and commitment but also the structure that dictates how we should work.* [Informant 6]

Adaptability to Workplace Change

Reorganizations that occurred at some of the centers were described as a major obstacle in the implementation of Sisom, as well as changes in the form of some professionals who were committed to working with Sisom had ended their employment or had gone on parental leave during the implementation. Such reorganizations resulted in both the physical and structural conditions to work with Sisom as intended were altered, which resulted in the work with Sisom being less prioritized or forgotten for a period. The loss of people who had been assigned specific roles in the implementation process made it difficult to maintain continuity in the process and left more work for those who remained. Despite this, the implementation of Sisom had a high priority among the professionals as it was perceived as valuable for the care process:

> *...The challenge is often the continuity of those people who have got an assignment that they are to stay in their jobs, many have finished. The staff turnover rate is high. Continuity is a challenge nowadays.* [Informant 10]

Implementation

A total of 3 subcategories emerged from the data: *structure, and content of the course*, *support from facilitators and researchers*, and *recruitment and reach*.

Structure and Content of the Course

Overall, the professionals had positive experiences of the content in the course. The sessions at the centers were carried out as planned. The professionals appreciated the dialogue during the session and found it valuable to reflect on what child participation means in general and how they currently worked with child participation at the center. It was especially important for new professionals, who did not have the same experience or knowledge about the importance of highlighting the child's perspective in their practical work. All the professionals stated that Sisom worked as a motivator for their ambition to improve child participation in care and that they were very enthusiastic about using it. They appreciated that they were given the opportunity to practice using Sisom by themselves for a few days before beginning to use it with the children. Adjustments were made at all centers regarding the implementation of the intervention to suit their routines and purpose:

> *It (the course) was good, it was good to get the opportunity to discuss participation and what it can mean and how we perceived it. The exercises and being able to test and navigate in the app were also very good. Then I can think that if we had had...another session after some of us had started because we discovered that new questions arose...mostly concerning the usage.* [Informant 3]

Support From Facilitators and Researchers

The facilitator's role differed somewhat among the included centers. At those where 4 or more professionals participated in the project, a facilitator was appointed to be responsible, motivate, follow up, and maintain overall contact with the researchers. At the centers with only 2 professionals working with Sisom, this facilitating role was not needed as they worked closely together and shared responsibility for the implementation process. Some of the first appointed facilitators were replaced by other people because of parental leave or termination of employment, which interrupted the implementation process.

In some cases, the facilitator did not work practically with Sisom. This was described as being both advantageous and disadvantageous. The advantage was that the facilitator had allocated time to be available to others and had an overview of the implementation process. The disadvantage was that the facilitator lacked personal experience from working practically with Sisom, which would have been valuable in supporting the professionals in their work. An important function of the

facilitator was to deal with and support the professionals regarding issues or problems during the implementation:

> *encourage to have...a coordinating meeting and when necessary to be able to have direct contact with the staff who work with Sisom if it is needed. I think that it makes a great difference that they can meet each other...to get together and think about what we need to be able to do this in the best possible way. That one is forced to think and not just solve.* [Informant 9]

Confidence and trust in the researchers were described as promoting factors for the feasibility of the implementation. The professionals experienced transparency, a clear common agenda as well as availability and continuous coordination as important factors that contributed to their trust in the project. The design and structure of information material, questionnaires, checklists, etc, were experienced by the professionals as simplifying their use. Confidence was also created by researchers being prepared to modify the plan for the intervention and the evaluation:

> *I felt that there was always the possibility to ask questions.* [Informant 7]

Recruitment and Reach

The recruitment of staff to the project was based on the inclusion of those showing interest in joining the project. A total of 17 professionals were included in the courses of whom 15 worked practically with Sisom. In addition to these, 3 people ended their involvement because of sick leave, parental leave, and termination of employment. A total of 46 children at baseline and 33 children at follow-up participated in the project (Table 3).

All the children who met the inclusion criteria were asked about participation. The reason why the professionals did not reach more children at baseline was because of a limited number of planned appointments during the intervention period. Some children were excluded based on them not understanding Swedish. Most children and parents agreed to participate, and the reason for not being willing to participate was of a practical nature, having to come to the hospital more often thus requiring more time away from work and school. The reason for the drop-out at the follow-up (n=13) was that professionals and children did not have the possibility to meet each other twice or that some of the children died during the project period.

The professionals anticipated that the age range for inclusion could be extended. Younger children (below 6 years) usually have sufficient digital experience and competence to use Sisom. Similarly, older children (over 12 years) could have the opportunity to decide if Sisom could be appropriate for them or not:

> *It's incredibly individual. We have thought that we have younger children who definitely could do this, and we have had those who are of an age that they are on the verge of being too old. Then, it's just a game. But even if it is then they sit down and discuss anyway. Then, it didn't become pointless anyway.* [Informant 12]

Table 3. Demographic characteristics of included children.

Characteristics	Baseline	Follow-up
Sites, n (%)		
University hospital	33 (72)	22 (67)
Rural hospital	13 (28)	11 (33)
Sex, n (%)		
Female	24 (52)	17 (51)
Age (years), n (%)		
6-9	21 (46)	12 (36)
10-13	24 (52)	21 (64)
Missing	1 (2)	0 (0)
Diagnosis, n (%)		
Hematological conditions	5 (11)	3 (9)
Cancer	15 (33)	12 (36)
HIV	7 (15)	4 (12)
Heart disease	4 (9)	4 (12)
Neurological conditions	8 (17)	7 (21)
Other	5 (11)	2 (6)
Missing	2 (4)	1 (3)

Mechanisms of Impact

A total of 2 subcategories emerged from the data, *applying Sisom in practice* and *technical and practical barriers.*

Applying Sisom in Practice

Sisom was mainly used when the children came on a planned follow-up appointment (30-60 min; see Multimedia Appendix 1). The professionals pointed out that it was important that the child used Sisom without being influenced by them or their parents. Some professionals and children wanted to discuss issues directly when Sisom was used. Most of them wrote the report and then started a discussion based on it, either during the same appointment or at a follow-up meeting a few weeks later. Some professionals asked the child and parents to discuss the report at home before the follow-up dialogue. The professionals usually chose to go through everything in the report from the child's first use of Sisom but focusing on things that were good or problematic:

> *...then we've looked and as I said we've been with the children and brought up positive things also because that's very important. And then we've talked about certain things where it doesn't look so good. Here's a red man, can you talk about it a little? And then we do that. It's much easier than getting a direct question. Then, they've been able to talk about how they have felt.* [Informant 12]

The children found it easier to have the written report or the tablet as support for the dialogue, helping them to focus on the report instead of the professional. All the children wanted to use Sisom, but some children did not want to talk about the report afterward. However, the staff felt that the child thought it was nice that they had the opportunity to talk about how they feel without having to sit for a long time and talk about their situation. Overall, the children were satisfied with the conversation based on Sisom and most of them wanted to show Sisom and the printed report to their parents. The report became their own product and some children felt proud of what they had done:

> *I think that many also feel it's good that mum and dad know that I actually think it's difficult or that I'm worried because they are worried. It becomes clearer. If one has said it once, then it's easier to say it again.* [Informant 11]

It was primarily the professionals who initiated what the discussion would focus on, and frequently, decisions were made based on the child's self-evaluations and described emotions in Sisom. The professionals pointed out that the child's reporting in Sisom cannot be interpreted as a measure of the child's feelings, but rather requires a subsequent conversation to hear how the child reasons about the output in the report and to understand what the child meant and the child's own view of their situation.

The professionals said that this way of working with Sisom requires preparation and planning on their part when it comes to charging the tablet, reserving a place where the child can use Sisom, printing the report, and then having time for a subsequent discussion.

Technical and Practical Barriers

The most common barriers during the use of Sisom were technical and related to the prototypic nature of the version of Sisom used for the implementation, especially related to the printing of reports, bugs in the system, and swift termination of one child's session before the onset of another. Bugs in the system were mostly related to the system freezing. Most often, children solved this by visiting another island and returning to the same question later on. However, the professionals felt that such technical disturbances disrupted the flow in the dialogue around the child's assessments and therefore need to be minimized. That one child could start using Sisom if the previous session had not been terminated correctly, resulted in uncertainties about the validity of the report. The professionals emphasized that the trustworthiness of the data is a key for acceptance and that the results in the printed reports must be unambiguous:

> *It took a while to print the reports, but no big problems, I don't think so, it was purely technical problem...yes, it was purely technical problem that I saw as possible problem, nothing else.* [Informant 7]

The professionals requested that Sisom should be available on the Web or for download so that the children could use Sisom at home before the meeting to increase effectiveness and to reduce waiting times at the clinic. Completing Sisom at home would also allow for more individualized use, based on individual capabilities and prerequisites.

Sisom was originally designed for children with cancer, and therefore some questions were less relevant for some children involved in this study. When a question was perceived as strange, the professionals recommended the child to skip it and move on. The children interpreted the questions based on their own situation, and the professionals thus stressed that it was important to not use the responses in the reports as absolute assessments of the children's life situation but rather as support for a later discussion with the child:

> *There were questions that I had not thought of but which were still important...but still, it concerns a lot of people. And just like cardiac patients, they often have this nausea and difficulties eating and how one sees oneself in various ways and how one thinks that one's friends see one. They're very important questions, and I don't think it's obvious to think of such questions as a professional when one works.* [Informant 8]

There is no possibility to modulate the use of Sisom, in its current form, such as pausing and saving the results to work with the questions at different times or submitting comments or notes while responding to questions or while reviewing the report. This was described as a hindering factor as children, who might not have time or energy to complete all islands/questions at once, could not pause their work and professionals could not write notes or recommendations for their own documentation or for forwarding information to a colleague. Furthermore, the professionals saw a need to establish guidelines for how and where the reports should be saved, which

actions should be taken based on specific information raised, and how to follow up on issues that were brought forward, and that was beyond the responsibility of that particular professional.

Experiences of Effects

A total of 2 subcategories emerged from the data: *improved participation of children* and *improved child-centered care*.

Improved Participation of Children

The major advantage of implementing Sisom was that it helped children express themselves independently without the pressure of an adult who directly lead or guided the dialogue. The professionals spoke of there being a culture today that the professionals primarily talk to the parents and when the children reach 18 years of age they are expected to be able to take full responsibility themselves. Sisom was described as being a useful tool to give the opportunity for the children to train and acquire ability to take responsibility for providing their information, to express themselves, to discuss their life situation, and to be able to participate in making decisions on issues that concern themselves, thus preparing them for adult life. With Sisom, the children asked more questions than usually and found it pleasant to put words on experiences that they had not previously dared to talk about.

Many children wanted to protect their parents and as a consequence, did not dare to ask questions or raise issues that concern them but that could worry their parents. This both relieved difficult feelings among the parents and made parents more aware of views about their child's experiences and feelings that they were not previously aware of:

> *It's a good way...because we, we also work with it, when they become older teenagers and will be going to the doctor's appointment by themselves, then it's good for both parents and the child to train themselves in that it's the child who is to talk, answer more and that the parents have to hold back and keep in the background a while during the appointment, and the child becomes more secure in talking with the doctor and us.* [Informant 14]

The professionals said that Sisom supported the children in comprehending and managing the context of the meeting with the professional resulting in that they both knew why and what they would talk about and that they were able to express themselves in a way that they felt comfortable with. The dedramatized questions in Sisom, which normally can be very sensitive or frightening, helped the children in processing their experiences about their life situation:

> *What he and his mother spoke of...is that they were very happy, satisfied, and positive. The boy said himself that he thought that it had been fun, but also that there were good questions, questions that made him think.* [Informant 5]

Similarly, the professional emphasized that the child became an actor and that the roles in their relationship clarified. In some cases, the format of the meeting was changed so that the child and the professional first met each other alone, and then, the parents were invited in. In this format, the child was involved in deciding what they wanted to talk about alone and together with their parents:

> *They participate more, just because it is their answers that are in focus during the appointment. They get more questions (from us). They are proud of their reports. They have been much more active in these appointments than they have been otherwise.* [Informant 14]

Improved Child-Centered Care

The experienced benefits of using Sisom were that the children's entire life situation was visualized, that the children's own experiences about their situation appeared, that changes over time could be captured, and that decisions could be made with the children or based on the children's perspectives on difficulties and needs. This led to increased participation of the children, and the professional's occupational roles changed. Sisom was described as a facilitator to support professionals in applying a child-centered care approach through these effects.

The staff talked of previously having focused on pathogenic perspectives, such as the diagnosis, symptoms, treatment, and hospital-related issues. By applying Sisom, they experienced an increased ability to understand the child's entire life situation. Working with Sisom also supported them in understanding the child's resources, strengths, and things that were positive in the child's situation. Sisom helped them in capturing and visualizing the child's own experiences and gave a clearer picture of the child's way of thinking and experiences of the situation. The effort and extra time spent on using Sisom were therefore experienced as an investment:

> *Because Sisom has a much broader approach when it also covers the home environment and how one feels at school. We are more, our material is more focused on the chronic illness itself and one's experiences being at the hospital.* [Informant 2]

Through the use of Sisom, the professionals also became aware of areas in the child's life that needed attention but that would not normally be addressed based on the responsibilities of their professional role. This broadened awareness of the experiences and needs of the child changed their professional role in a way that shifted their perspective from being grounded in their professional expertise toward what the particular child communicated. They experienced that they would never have been able to ask all the questions provided in Sisom in a regular conversation, as it would have been both boring and tiresome for the child. The broad awareness of the child's life situation provided by Sisom helped the professionals in not fractioning the child's needs into the responsibilities of separate professional roles. This gave them a broader view and a new way of thinking about their professional role:

> *...We start talking about essential matters and it's not because I'm a difficult social worker who asks questions about that sort of thing, but it's the game that leads us on to these essential things...it's not me who's completing some paper but it's something the child does and there are no secrets, these aren't things I hide and don't show, and then the child can take it*

> *home and look at it and talk with mum and dad...* [Informant 7]

On the same note, Sisom changed the spirit of the dialogue between the professional and the child, from the form of a structured questioning to a dialogue driven by curiosity. The professionals traditionally have all the power and lead the dialogue, thus reinforcing their dominant role and the role of the child as a subject being interrogated. Through the use of Sisom, they did not have to lean on their own or the parent's interpretations of the children's experiences and needs. Instead, Sisom helped them to provide information about completely different things—things that were important to the child and things that neither the professionals nor the parents knew. This made the professionals feel more relaxed in their dialogue with the child as they felt that Sisom had caught the children's perspective on what was needed to be discussed. This placed the conversation with the child at a deeper level, creating more space and time to talk about other things than they usually do. The professionals reflected that they now could approach the child in a different way, allowing the child to talk about the situation in their own way and thereby getting the *true* information from the child. Above all, Sisom was especially helpful in supporting ways of talking with those children who the professionals usually have difficulties in talking to and those children who only answer *yes* or *no* to direct questions. In these cases, Sisom created a distance to the actual questions, and the professional could act more like a sidekick in the *conversation* provided by the child's interaction with Sisom. The professional's role became less inquisitive and nosey, making the children more involved and share more information about their life situation:

> *There were questions that at first sight didn't feel relevant, but when we spoke more about the situation it emerged that it was relevant for the boy. For example, questions about death. It isn't what we primarily work with in relation to diabetes. There are no children who die because of diabetes. But he had questions and thoughts about it, and so we could talk more about it later. Most importantly, his mother got the chance to follow up his thoughts at home.* [Informant 5]

The use of Sisom also increased the professional's ability to detect changes in the child's situation over time. When they and the child compared the latter's assessments in Sisom at different time points, the children were sometimes surprised over the change that had occurred. The professionals also said that the actions decided on from their discussion with the child were more person-centered because of improved awareness of the child's situation and needs.

Discussion

Principal Findings

In this study, we used a process evaluation design [38] to study the implementation of the eHealth service Sisom at 4 pediatric care centers in Sweden. The weak evidence base for eHealth services in general can delay their implementation in practice [37]. When developing eHealth services and planning for evaluation, ongoing process evaluations are not commonly used. Furthermore, process evaluations per se are not commonly including both design and pretesting phases and evaluation of the clinical, human, social, and organizational aspects of the eHealth services following implementation [37]. Thus, this implementation study fills a knowledge gap by investigating facilitators and barriers that are of particular importance for successful implementation of an eHealth solution, particularly in the context of children's participation in pediatric care. This study specifically contributes to a more comprehensive evidence base for the eHealth service Sisom. Previous research on Sisom has primarily focused on the design and pretesting phases whereas this study is the first one to investigate the implementation of Sisom in clinical practice. Further studies of implementation and effects will be needed as the use of Sisom is scaled up, reaching a more diverse array of users and settings for implementation.

The data presented in this study provided support for the strategy applied for the implementation of Sisom at the pediatric care centers involved. The professionals were satisfied with the anchoring of the project, with the course offered, and with the support provided by the local facilitators and the researchers during the project. Overall, they experienced that each of these aspects was supportive for introducing Sisom in daily practice. In changing and restructuring the way of working with patient participation in practice, it is crucial to consider contextual factors such as culture, resources, and priorities [42]. However, focusing only on the changing of knowledge, attitudes, and behavior of the professionals does not suffice. This is confirmed by professionals who often feel left alone in their effort to facilitate children's participation [43]. In our study, the shared commitment and value base toward child-centered care at the health care centers were described as facilitating factors for a positive effective response to the intervention. Previous research has established that interventions that fit well with the organization's existing culture and values are more likely to be successful [42]. The professionals involved in the introduction of Sisom emphasized that the knowledge base and underlying principles of the concept of participation discussed during the education helped them to generate enthusiasm for the intervention and to promote their capacity to successfully implement Sisom in practice at their center. Thus, the contextual factors affecting successful implementation were of a type that is traditionally expected to be hard to change in health care settings such as the values and attitudes, readiness for behavioral change, and the added value of the intervention [38,44,45]. Another facilitating factor for the implementation of the intervention was that the children, who were the intended users of Sisom, accepted the design of the solution and experienced meaningfulness in its use, which can probably be explained by the high degree of child participation in the development of Sisom [29]. However, there were a few contextual obstacles that could be important to consider for achieving successful implementation, such as workplace reorganization, changes in employment, changes in priorities for development work, and time constraints. Interesting findings in relation to this were that it was easier to implement Sisom at a center where professionals worked independently in relation to their colleagues, managing their own time and meeting the children

individually. As independent professionals, they were able to plan and restructure their scheduling of meetings based on the children's needs, individually planning the length of meetings and deciding to invite children to extra meetings for going through the reports from Sisom.

This implementation study provides several practical examples of how Sisom contributes to improvements in child participation. This is significant as most efforts to implement a child-centered perspective in pediatric health care have little or no evidence that those practices really provide opportunities for children to share their views, needs, and preferences and to participate in conversations about their own care [7,46,47]. The major benefit with the implemented intervention was that by using Sisom, the children could express themselves independently without the pressure from an adult and without the risk of being dominated by an adult. The professionals testified that it became evident that Sisom helped the children to dare to ask questions and talk about their experiences, thoughts, and feelings, and that the purpose of the dialogue with the health care professional became less vague and more relevant. The professionals experienced that Sisom increased their ability to detect changes in the child's situation over time and to strengthen the children in being an actor in their own health care. Sisom also opened up different and further conversations within the family, and parents became more aware of their child's experiences and feelings. Sisom also had an effect on the role of the parents in conversations with the professionals: from being someone who had a responsibility for responding on behalf of their child, to someone who actively listened to their child and provided support in communicating what was important to them in relation to their life situation.

Limitation

Some limitations should be considered when interpreting the results. First, the recruitment of professionals to the project was limited to those having an interest in participating in the implementation. Participation based on voluntariness means that perspectives and experiences from professionals who would have been forced to take part in the implementation of Sisom are thus lacking. This setting limited the general impact the implementation had on the whole workplace and also made the project more vulnerable to changes in priorities and reorganizations at the participating centers. Second, further studies are needed in other health care settings and in contextual conditions in health care systems in other countries to strengthen the conclusions on the validity and generalizability of the implementation strategy applied here. Third, the implementation of Sisom needs to be investigated both longitudinally and in randomized controlled studies to analyze its long-term effects on organizations, professional roles, ways of working, and ultimately on children's health outcomes.

Conclusions

The changing nature of health care delivery from a provider-centered approach to an increasingly child-centered approach in which the children become actors in their own care is a challenging process. We believe this study represents a significant contribution to this field of research. To our knowledge, this is the first implementation study of an eHealth service aimed at strengthening children's participation in pediatric health care. The results from the study show that it is possible to restructure health care delivery toward a child-centered approach if there is a willingness and preparedness in the organization to implement an eHealth solution with the aim of restructuring the way of working with children's participation. Key factors for successful implementation are alignment of the solution with the values and goals of the organization, health care professionals' beliefs in the usefulness and usability of the solution, and health care professionals' willingness to change their professional roles promoted by the solution.

Acknowledgments

The authors would like to thank all the participants who participated in this study and who made this study possible. This study was funded by the Swedish Childhood Cancer Foundation, The Swedish Rheumatism Association, and the Center of Research on Welfare, Health, and Sport at Halmstad University. The authors would like to acknowledge Professor Cornelia Ruland and the Center for Shared Decision Making and Collaborative Care Research at Norway University Hospital for valuable support in making Sisom available to the centers participating in this study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Description of how Sisom was applied, at baseline and at follow-up.
[PDF File (Adobe PDF File), 131 KB - jmir_v21i11e14271_app1.pdf]

References

1. The Office of the United Nations High Commissioner for Human Rights. 1989. Convention on the Rights of the Child URL: https://www.ohchr.org/documents/professionalinterest/crc.pdf [accessed 2019-08-22]
2. Riksdagens Öppna Data. 2017. Regeringens Proposition 2017/18:186 Inkorporering Av FN:s Konvention Om Barnets Rättigheter URL: https://data.riksdagen.se/fil/F99B084B-5202-4EDD-9E5D-D3B3D44DA1E9 [accessed 2019-08-22]

3. Riksdagens Öppna Data. 2014. Patientlag (2014:821) URL: https://www.riksdagen.se/sv/dokument-lagar/dokument/svensk-forfattningssamling/patientlag-2014821_sfs-2014-821 [accessed 2019-08-22]
4. van Bijleveld GG, Dedding CW, Bunders-Aelen JF. Children's and young people's participation within child welfare and child protection services: a state-of-the-art review. Child Fam Soc Work 2013 Jul 4;20(2):129-138. [doi: 10.1111/cfs.12082]
5. Imms C, Adair B, Keen D, Ullenhag A, Rosenbaum P, Granlund M. 'Participation': a systematic review of language, definitions, and constructs used in intervention research with children with disabilities. Dev Med Child Neurol 2016 Jan;58(1):29-38 [FREE Full text] [doi: 10.1111/dmcn.12932] [Medline: 26411643]
6. Larsson I, Staland-Nyman C, Svedberg P, Nygren JM, Carlsson IM. Children and young people's participation in developing interventions in health and well-being: a scoping review. BMC Health Serv Res 2018 Jun 28;18(1):507 [FREE Full text] [doi: 10.1186/s12913-018-3219-2] [Medline: 29954392]
7. Coyne I, Amory A, Kiernan G, Gibson F. Children's participation in shared decision-making: children, adolescents, parents and healthcare professionals' perspectives and experiences. Eur J Oncol Nurs 2014 Jun;18(3):273-280. [doi: 10.1016/j.ejon.2014.01.006] [Medline: 24629505]
8. Feenstra B, Boland L, Lawson ML, Harrison D, Kryworuchko J, Leblanc M, et al. Interventions to support children's engagement in health-related decisions: a systematic review. BMC Pediatr 2014 Apr 23;14:109 [FREE Full text] [doi: 10.1186/1471-2431-14-109] [Medline: 24758566]
9. Gilljam BM, Arvidsson S, Nygren JM, Svedberg P. Promoting participation in healthcare situations for children with JIA: a grounded theory study. Int J Qual Stud Health Well-being 2016;11:30518 [FREE Full text] [doi: 10.3402/qhw.v11.30518] [Medline: 27172512]
10. Moore L, Kirk S. A literature review of children's and young people's participation in decisions relating to health care. J Clin Nurs 2010 Aug;19(15-16):2215-2225. [doi: 10.1111/j.1365-2702.2009.03161.x] [Medline: 20659201]
11. Sjöberg C, Amhliden H, Nygren JM, Arvidsson S, Svedberg P. The perspective of children on factors influencing their participation in perioperative care. J Clin Nurs 2015 Oct;24(19-20):2945-2953. [doi: 10.1111/jocn.12911] [Medline: 26215896]
12. Virkki M, Tolonen TH, Koskimaa T, Paavilainen E. Children as decision-makers in health care - an integrative review. Clin Nurs Stud 2015;3(1):47-54. [doi: 10.5430/cns.v3n1p47]
13. Wyatt KD, List B, Brinkman WB, Lopez GP, Asi N, Erwin P, et al. Shared decision making in pediatrics: a systematic review and meta-analysis. Acad Pediatr 2015;15(6):573-583. [doi: 10.1016/j.acap.2015.03.011] [Medline: 25983006]
14. Coyne I, Amory A, Kiernan G, Gibson F. Children's participation in shared decision-making: children, adolescents, parents and healthcare professionals' perspectives and experiences. Eur J Oncol Nurs 2014 Jun;18(3):273-280. [doi: 10.1016/j.ejon.2014.01.006] [Medline: 24629505]
15. Vatne TM, Finset A, Ørnes K, Holmstrøm H, Småstuen MC, Ruland CM. Effects of an interactive symptom communication tool for children with heart disease on patient-provider communication in outpatient care: preliminary results. J Commun Healthc 2013 Nov 15;6(2):106-114. [doi: 10.1179/1753807612y.0000000020]
16. Ruland CM, Starren J, Vatne TM. Participatory design with children in the development of a support system for patient-centered care in pediatric oncology. J Biomed Inform 2008 Aug;41(4):624-635 [FREE Full text] [doi: 10.1016/j.jbi.2007.10.004] [Medline: 18082468]
17. Adair B, Ullenhag A, Keen D, Granlund M, Imms C. The effect of interventions aimed at improving participation outcomes for children with disabilities: a systematic review. Dev Med Child Neurol 2015 Dec;57(12):1093-1104 [FREE Full text] [doi: 10.1111/dmcn.12809] [Medline: 26010935]
18. Gehring ND, McGrath P, Wozney L, Soleimani A, Bennett K, Hartling L, et al. Pediatric e-mental healthcare technologies: a systematic review of implementation foci in research studies, and government and organizational documents. Implement Sci 2017 Jun 21;12(1):76 [FREE Full text] [doi: 10.1186/s13012-017-0608-6] [Medline: 28637479]
19. Grist R, Porter J, Stallard P. Mental health mobile apps for preadolescents and adolescents: a systematic review. J Med Internet Res 2017 May 25;19(5):e176 [FREE Full text] [doi: 10.2196/jmir.7332] [Medline: 28546138]
20. Wozney L, McGrath PJ, Gehring ND, Bennett K, Huguet A, Hartling L, et al. eMental healthcare technologies for anxiety and depression in childhood and adolescence: systematic review of studies reporting implementation outcomes. JMIR Ment Health 2018 Jun 26;5(2):e48 [FREE Full text] [doi: 10.2196/mental.9655] [Medline: 29945858]
21. Badawy SM, Cronin RM, Hankins J, Crosby L, DeBaun M, Thompson AA, et al. Patient-centered ehealth interventions for children, adolescents, and adults with sickle cell disease: systematic review. J Med Internet Res 2018 Jul 19;20(7):e10940 [FREE Full text] [doi: 10.2196/10940] [Medline: 30026178]
22. Mehdizadeh H, Asadi F, Mehrvar A, Nazemi E, Emami H. Smartphone apps to help children and adolescents with cancer and their families: a scoping review. Acta Oncol 2019 Jul;58(7):1003-1014. [doi: 10.1080/0284186X.2019.1588474] [Medline: 30915872]
23. Raaff C, Glazebrook C, Wharrad H. A systematic review of interactive multimedia interventions to promote children's communication with health professionals: implications for communicating with overweight children. BMC Med Inform Decis Mak 2014 Jan 22;14:8 [FREE Full text] [doi: 10.1186/1472-6947-14-8] [Medline: 24447844]
24. Unni E, Gabriel S, Ariely R. A review of the use and effectiveness of digital health technologies in patients with asthma. Ann Allergy Asthma Immunol 2018 Dec;121(6):680-91.e1. [doi: 10.1016/j.anai.2018.10.016] [Medline: 30352288]

25. Livingstone S, Lemish D, Lim SS, Bulger M, Cabello P, Claro M, et al. Global perspectives on children's digital opportunities: an emerging research and policy agenda. Pediatrics 2017 Nov;140(Suppl 2):S137-S141 [FREE Full text] [doi: 10.1542/peds.2016-1758S] [Medline: 29093049]
26. Veneri DA, Zdanis K. Are technology-informed methods better than traditional approaches in educating patients? A meta-analysis. Health Educ J 2018 Jan 9;77(3):261-276. [doi: 10.1177/0017896917741511]
27. Garell C, Svedberg P, Nygren JM. A legal framework to support development and assessment of digital health services. JMIR Med Inform 2016 May 25;4(2):e17 [FREE Full text] [doi: 10.2196/medinform.5401] [Medline: 27226391]
28. Ruland CM, Slaughter L, Starren J, Vatne TM, Moe EY. Children's contributions to designing a communication tool for children with cancer. Stud Health Technol Inform 2007;129(Pt 2):977-982. [Medline: 17911861]
29. Arvidsson S, Gilljam B, Nygren J, Ruland CM, Nordby-Bøe T, Svedberg P. Redesign and validation of Sisom, an interactive assessment and communication tool for children with cancer. JMIR Mhealth Uhealth 2016 Jun 24;4(2):e76 [FREE Full text] [doi: 10.2196/mhealth.5715] [Medline: 27343004]
30. Baggott C, Baird J, Hinds P, Ruland CM, Miaskowski C. Evaluation of Sisom: a computer-based animated tool to elicit symptoms and psychosocial concerns from children with cancer. Eur J Oncol Nurs 2015 Aug;19(4):359-369. [doi: 10.1016/j.ejon.2015.01.006] [Medline: 25691300]
31. Siedlikowski M, Rauch F, Ruland C, Tsimicalis A. Giving children a voice: development of an interactive assessment and communication tool for osteogenesis imperfecta. J Pain Symptom Manag 2018 Dec;56(6):e103. [doi: 10.1016/j.jpainsymman.2018.10.348]
32. Tsimicalis A, le May S, Stinson J, Rennick J, Vachon M, Louli J, et al. Linguistic validation of an interactive communication tool to help French-speaking children express their cancer symptoms. J Pediatr Oncol Nurs 2017;34(2):98-105. [doi: 10.1177/1043454216646532] [Medline: 27226380]
33. Tsimicalis A, Rennick J, le May S, Stinson J, Sarkis B, Séguin K, et al. 'Tell it as it is': how Sisom prompts children and parents to discuss their cancer experience. Cancer Rep 2019 Apr 23:e1173 (epub ahead of print)(forthcoming)(forthcoming). [doi: 10.1002/cnr2.1173]
34. Tsimicalis A, Stone PW, Bakken S, Yoon S, Sands S, Porter R, et al. Usability testing of a computerized communication tool in a diverse urban pediatric population. Cancer Nurs 2014;37(6):E25-E34 [FREE Full text] [doi: 10.1097/NCC.0000000000000115] [Medline: 24457227]
35. Vatne TM, Slaugther L, Ruland CM. How children with cancer communicate and think about symptoms. J Pediatr Oncol Nurs 2010;27(1):24-32. [doi: 10.1177/1043454209349358] [Medline: 19833978]
36. Pfadenhauer LM, Gerhardus A, Mozygemba K, Lysdahl KB, Booth A, Hofmann B, et al. Making sense of complexity in context and implementation: the context and implementation of complex interventions (CICI) framework. Implement Sci 2017 Feb 15;12(1):21 [FREE Full text] [doi: 10.1186/s13012-017-0552-5] [Medline: 28202031]
37. Enam A, Torres-Bonilla J, Eriksson H. Evidence-based evaluation of ehealth interventions: systematic literature review. J Med Internet Res 2018 Nov 23;20(11):e10971 [FREE Full text] [doi: 10.2196/10971] [Medline: 30470678]
38. Moore GF, Audrey S, Barker M, Bond L, Bonell C, Hardeman W, et al. Process evaluation of complex interventions: Medical Research Council guidance. Br Med J 2015 Mar 19;350:h1258 [FREE Full text] [doi: 10.1136/bmj.h1258] [Medline: 25791983]
39. Eriksson L, Huy TQ, Duc DM, Selling KE, Hoa DP, Thuy NT, et al. Process evaluation of a knowledge translation intervention using facilitation of local stakeholder groups to improve neonatal survival in the Quang Ninh province, Vietnam. Trials 2016 Jan 13;17:23 [FREE Full text] [doi: 10.1186/s13063-015-1141-z] [Medline: 26762125]
40. Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
41. WMA – The World Medical Association. 2013. WMA Declaration of Helsinki – Ethical Principles for Medical Research Involving Human Subjects URL: https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/ [accessed 2019-08-22]
42. Scholl I, LaRussa A, Hahlweg P, Kobrin S, Elwyn G. Organizational- and system-level characteristics that influence implementation of shared decision-making and strategies to address them - a scoping review. Implement Sci 2018 Mar 9;13(1):40 [FREE Full text] [doi: 10.1186/s13012-018-0731-z] [Medline: 29523167]
43. Carlsson IM, Nygren JM, Svedberg P. Patient participation, a prerequisite for care: a grounded theory study of healthcare professionals' perceptions of what participation means in a paediatric care context. Nurs Open 2018 Jan;5(1):45-52 [FREE Full text] [doi: 10.1002/nop2.106] [Medline: 29344394]
44. Aarons GA, Green AE, Trott E, Willging CE, Torres EM, Ehrhart MG, et al. The roles of system and organizational leadership in system-wide evidence-based intervention sustainment: a mixed-method study. Adm Policy Ment Health 2016 Nov;43(6):991-1008 [FREE Full text] [doi: 10.1007/s10488-016-0751-4] [Medline: 27439504]
45. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implement Sci 2009 Aug 7;4:50 [FREE Full text] [doi: 10.1186/1748-5908-4-50] [Medline: 19664226]
46. Coyne I, Gallagher P. Participation in communication and decision-making: children and young people's experiences in a hospital setting. J Clin Nurs 2011 Aug;20(15-16):2334-2343. [doi: 10.1111/j.1365-2702.2010.03582.x] [Medline: 21410571]

47. Koller D. 'Kids need to talk too': inclusive practices for children's healthcare education and participation. J Clin Nurs 2017 Sep;26(17-18):2657-2668. [doi: 10.1111/jocn.13703] [Medline: 28001330]

Abbreviations

eHealth: electronic health
UNCRC: United Nation’s Convention of the Rights of the Child

Edited by G Eysenbach; submitted 08.04.19; peer-reviewed by J Farzi, H Ebrahimpour Sadagheyani; comments to author 19.05.19; revised version received 04.07.19; accepted 19.07.19; published 15.11.19.

Please cite as:
Svedberg P, Arvidsson S, Larsson I, Carlsson IM, Nygren JM
Barriers and Enablers Affecting Successful Implementation of the Electronic Health Service Sisom: Multicenter Study of Child Participation in Pediatric Care
J Med Internet Res 2019;21(11):e14271
URL: https://www.jmir.org/2019/11/e14271
doi:10.2196/14271
PMID:31730040

Original Paper

Health-Related Internet Use Among Men With Prostate Cancer in Canada: Cancer Registry Survey Study

Jacqueline L Bender[1,2,3], PhD; Deb Feldman-Stewart[4,5], PhD; Christine Tong[5], MA; Karen Lee[1], MScPT; Michael Brundage[4,5], MD, MSc; Howard Pai[6,7], MD; John Robinson[8,9], PhD; Tony Panzarella[2], MSc

[1]ELLICSR Cancer Rehabilitation and Survivorship Program, Department of Supportive Care, Princess Margaret Cancer Centre, Toronto, ON, Canada
[2]Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada
[3]Institute of Health Policy, Management, and Evaluation, University of Toronto, Toronto, ON, Canada
[4]Department of Oncology, Queen's University, Kingston, ON, Canada
[5]Division of Cancer Care and Epidemiology, Queen's Cancer Research Institute, Kingston, ON, Canada
[6]Division of Radiation Oncology, BC Cancer, Victoria, BC, Canada
[7]Department of Surgery, University of British Columbia, Vancouver, BC, Canada
[8]Department of Psychosocial and Rehabilitation Oncology, Tom Baker Cancer Centre, Calgary, AB, Canada
[9]Department of Oncology, University of Calgary, Calgary, AB, Canada

Corresponding Author:
Jacqueline L Bender, PhD
ELLICSR Cancer Rehabilitation and Survivorship Program
Department of Supportive Care
Princess Margaret Cancer Centre
Munk Building, B PMB 130
585 University Avenue
Toronto, ON, M5G 2C4
Canada
Phone: 1 416 581 8606
Email: jackie.bender@uhnresearch.ca

Abstract

Background: After a prostate cancer diagnosis, men want information about their disease and treatment options. The internet offers a convenient means to deliver health information to patients with prostate cancer. However, there are concerns about the use of the internet among this largely senior population.

Objective: This study aimed to determine the patterns and factors associated with the use of the internet as a source of health information among Canadian men with prostate cancer and the features and information required in a website.

Methods: Population surveys were conducted in four Canadian provinces (British Columbia, Alberta, Saskatchewan, and Ontario) in 2014-2015. Data analyses included descriptive, bivariable, and multivariable analyses. The Pearson Chi-square and univariable regression were used to examine associations between independent variables and health-related internet use. Correlates of health-related internet use were analyzed using multivariable logistic regression.

Results: A total of 1362 patients responded across the four provinces. The mean age of respondents was 69 years (SD 8.2). In addition, 82% (n=1071) were internet users and 71% (n=910) used the internet daily. Further, 65% (n=784) used the internet as a source of prostate cancer information, and 40% (n=521) were confident about using information obtained from the internet to make health decisions. Men who used the internet to obtain prostate cancer information were more likely to be active information seekers (odds ratio [OR]: 4.5, 95% CI 2.6-7.8), be confident using information from the internet to make health decisions (OR: 3.6, 95% CI 2.3-5.7), have broadband internet access (OR: 1.8, 95% CI 1.2-2.7), and have more unmet supportive care needs (OR: 1.05, 95% CI 1.0-1.1). Top features wanted in a website, reported by more than 50% of respondents, were a library of resources (n=893, 65.6%), tools to support treatment decision making (n=815, 59.8%), and tools to help navigate the prostate cancer journey (n=698, 51.2%). Top three topics of information wanted in such a website were treatment options (n=916, 67.3%), disease progression (n=904, 66.4%), and management of side effects (n=858, 63%).

Conclusions: Over two-thirds of Canadian patients with prostate cancer surveyed use the internet as a source of health information about prostate cancer, but over half did not feel confident using information from the internet to make health decisions. Being an

active information seeker, having confidence in using information from the internet to make health decisions, having broadband internet, and having more unmet supportive care needs were significantly associated with health-related internet use. Future work should examine electronic health literacy interventions as a means to boost men's confidence in using information from the internet and design websites that include information and features that help men navigate the prostate cancer journey and support treatment decision making and management of side effects.

(J Med Internet Res 2019;21(11):e14241) doi:10.2196/14241

KEYWORDS

prostate cancer; internet; health decision making; digital divide

Introduction

An estimated 1.3 million men worldwide were diagnosed with prostate cancer in 2018 [1] and 70% were from developed countries [2]. In Canada, one in seven men will be diagnosed with prostate cancer during their lifetime [3]. The risk of prostate cancer increases with age, such that 40% of all prostate cancer cases occur in men aged 60-69 years [3]. With the aging baby boom population, the number of Canadian men diagnosed with prostate cancer per year is expected to reach 42,000 by 2030 [3].

After a prostate cancer diagnosis, men want information about their disease and treatment options [4]. Information can enhance understanding, correct misconceptions about treatment, assist in coping with illness, engender feelings of control, and reduce anxiety by enabling patients to prepare for and predict aversive events [5,6]. Previous research has shown that the information needs of patients with prostate cancer are similar across time and different developed countries [4]. However, the specific amount and details of information needed varies, as do the reasons for wanting information [7].

The internet offers a convenient and cost-efficient way to provide personally tailored health information and services [8,9] and is expected to help reduce social inequalities by providing individuals access to information that might otherwise be inaccessible [10]. However, there is concern that the internet may be perpetuating disparities in access to health services by keeping certain segments of the population, such as seniors, on the sidelines [11]. According to the 2016 General Social Survey by Statistics Canada, a digital divide based on age still exists in Canada, with internet use being the lowest among those aged ≥75 years [12]. At the same time, Canadians aged 65-74 are the fastest-growing segment of internet users, with 81% using the internet in 2016 compared to 65% in 2013 [12].

Previously, we reported that for men with prostate cancer in Canada, the most frequently preferred sources of information about prostate cancer and its treatment are a urologist (96%), followed by a family doctor (90%), printed information (85%), other cancer patients (69%), and the internet (68%) [13]. These findings reflect a five-fold increase in the prevalence of health-related internet use among men with prostate in Canada compared to 10 years ago when only 12% of Canadian patients with prostate cancer used the internet as a health resource [14]. Moreover, recent data from the US Health Information National Trends survey (HINTs) shows an increasing trend toward using the internet as the first source of health information compared to family/friends/coworkers, health care professionals, and traditional media [10].

Although other studies have examined online health information–seeking behavior in the general population [10,11] and cancer survivors specifically [15], there are no recent studies on the health-related internet use of patients with prostate cancer. Current population-based data describing patterns of health-related internet use among patients with prostate cancer, as well as their confidence in using that information for making health decisions, could inform patient education and health service efforts. At the same time, understanding the factors associated with health-related internet use would be useful in designing strategies for reducing the digital divide [16,17]. In this paper, we report on the patterns and factors associated with the use of the internet as a source of health information among Canadian men with prostate cancer and the information and features wanted in a website.

Methods

Study Design

We conducted a cross-sectional survey using a modified Dillman [18] survey methodology, following the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) reporting standards for cross-sectional surveys [19]. Survey packages included an addressed, stamped, return envelope, which was to be returned after the survey was completed. After 4 weeks, a second survey package was sent to nonrespondents.

Setting and Participants

We surveyed men diagnosed with prostate cancer in four Canadian provinces—British Columbia, Alberta, Saskatchewan, and Ontario—in 2014-2015 using their respective provincial cancer registries. We sought to obtain responses from 10% of the provincial patients and expected a 30% response rate. Thus, to achieve responses from 10% of the target population in each province, each registry invited a random selection of 55%-60% of men diagnosed with prostate cancer in the last 6 months of 2012 to participate in the study. We selected 2012, as it was the latest year for which the registries had complete data. The only inclusion criteria were that the patient had a prostate cancer diagnosis and lived in that province that year.

Three registries (British Columbia, Alberta, and Saskatchewan) used an "opt-out" recruitment strategy, whereby the registry provided a cover letter introducing the study in the survey package, making clear that the recipient could choose whether

to complete the survey. The fourth registry (Ontario) used an "opt-in" recruitment strategy, providing a letter introducing the study and required the recipient to phone the registry to express their interest in participating. The names of interested patients from the opt-out province were then forwarded to the Ontario Lead at the study central office from where the survey was mailed out.

Each survey had a unique study identification code and did not include any identifying information. For the opt-out provinces, the provincial registries identified nonresponders and sent out the second survey 4 weeks later. For the opt-in province, the Ontario lead identified nonresponders and sent out the second survey 4 weeks later. All completed surveys were eventually mailed to the Ontario Lead at the study central office to be entered in an electronic database for analysis. Ethical approval was attained from each province's respective university and cancer agency and from the Ontario lead's university.

Questionnaire

The questionnaire contained five sections with 40 questions. It was developed by a team of researchers and health care professionals and piloted with five patients prior to implementation [13]. This paper reports on questions pertaining to patterns of internet access and use (n=7), confidence using the internet as a source of health information (n=1), factors associated with health-related internet use including participant demographics (n=8), and what men with prostate cancer want in a website (n=2). All responses were self-reported.

Questions pertaining to internet access and use included use of the internet (yes/no), type of internet access (dial-up, broadband, cellular network, wireless network, or not sure), devices used to access the internet (desktop/laptop computer, tablet, cellphone/smartphone, or other), internet access from home (yes/no), frequency of internet access (at least once a day, at least once a week, at least once a month, or less than once a month), and websites used for information on prostate cancer or its treatments (open response option).

Health-related internet use was measured by asking respondents to indicate "how easy or difficult it was to obtain information about prostate cancer and/or its treatments from each of the sources below" of which, "internet (other than personal email or online support groups)" was an option. Response categories included very easy, easy, difficult, very difficult (which were all coded as "yes"), and did not try to use this source (which was coded as "no").

Confidence in using the internet as a health resource was measured with one item from the eHealth Literacy Scale (eHEALS). The eHEALS is a validated measure of electronic health (eHealth) literacy, which is intended to capture people's perceived skills at using information technology for health [20]. Specifically, we adapted the final question in the eHEALS scale for prostate cancer: "I feel confident in using information from the Internet to help make health decisions [related to my prostate cancer]." We used the same response options as the eHEALS scale: a 5-point Likert scale (1=strongly disagree, 5=strongly agree) with the addition of, "Not applicable, I do not use information from the Internet to help make decisions related to my prostate cancer."

Lastly, preferences for using the internet as a health resource were investigated by asking participants to select information and features wanted in a website, from a list.

Variables

We examined associations between several variables and health-related information use. *Sociodemographic variables* included age (43-65, 66-75, ≥76 years), education completed (primary or secondary school/college or university), income (≤CAD 40,000, CAD 40,001-80,000, ≥CAD 80,001), area of residence (urban or suburban/rural, town or country), and broadband internet use (yes/no). *Internet experience* variables included frequency of internet use (every day/not every day), number of devices used to access the internet (one device/more than one device), and confidence in using information from the internet to inform health decisions (as described above). *Clinical variables* included treatments (active/active surveillance or watchful waiting). *Information seeking and decision making* variables included an information-seeking role (I did all the looking, I did some of the looking along with someone else/someone else did most or all of the looking, or I did want any information), and role in the decision making process (I made the decision by myself or with my doctor, family/my doctor or a family member made the decision). *Health status variables* included overall health (very good or good/poor or very poor) and number of unmet supportive care needs, as measured by the validated 34-item Supportive Care Needs Survey and the 8-item prostate cancer–specific module [21]. Responses were no need or not applicable, yes and the need is met (met need), and yes but the need is not met (unmet need). Unmet need items were summed to create a total unmet need score.

Data Analysis

Statistical Analysis

First, we carried out a descriptive analysis. Sample means and SDs were calculated for discrete/continuous variables and proportions for nominal variables. Chi-square tests and univariable logistic regression analyses were performed to test for associations between health-related internet use and variables hypothesized to be associated with such use. We assessed the construct validity of the eHealth confidence variable by examining correlations with other measures (eg, known components of eHealth literacy) in theoretically predictable ways using the Pearson Chi-square test [22]. Lastly, a multivariable logistic regression model was constructed to assess the relative importance of variables that were significantly associated (P<.05) with health-related internet use. Variables included in the final model were minimized through an iterative process of adding and removing variables until the model was considered to have a good fit [23]. Fit was determined based on significance values, and noting the presence or absence of interactions between variables [23].

Content Analysis of Open-Ended Responses

Responses to the open-ended question, "Are there any specific Internet site(s) that you like to go to for information related to prostate cancer or its treatments? If Yes, please specify," came in the form of bulleted phrases with examples. Hence, we used content analysis to code these data into explicit categories, which we then described using statistics [24]. One coder (KL) independently reviewed and itemized all unique responses (some participants provided more than one response). A codebook was developed and a coding strategy was applied to code each unique response and organize the responses into explicit categories (eg, national cancer agencies, hospital affiliated websites, commercial websites, and online patient communities). The coding strategy was refined following discussion with the lead author (JLB). A frequency count was performed to quantify the number and proportion of responses in each category.

Results

Sample Characteristics

A total of 1362 patients returned partially or fully completed surveys across provinces and 46 returned blank surveys. Blank surveys and missing responses from partially completed surveys were excluded from the analysis. The survey response rate for the opt-out provinces was 46%-55% and for the opt-in province, it was 13%. Table 1 shows respondents' sociodemographic, treatment, and general health characteristics. The age range of respondents was 43-95 years. Most were on follow-up after treatment (63%) and in good or very good health (93.6%).

Table 1. Respondent characteristics.

Characteristic	Count
Age (n=1320), mean (SD)	69.5 (8.2)
Relationship status (n=1313), n (%)	
With partner	1134 (86)
Without partner	179 (14)
Sexual orientation (n=1362), n (%)	
Heterosexual	1201 (88.2)
Gay	17 (1.2)
Bisexual	8 (0.6)
Education (highest level completed) (n=1312), n (%)	
Primary or secondary school	439 (33)
College, technical school, or university	873 (67)
Residence (n=1320), n (%)	
Urban or suburban	833 (63)
Rural, town, or country	487 (37)
Household income (n=1215), n (%)	
<CAD 40,000	347 (29)
CAD 40,001-80,000	454 (37)
>CAD 80,000	414 (34)
Language preference for health information (n=1302), n (%)	
English	1257 (96.5)
English and other language	18 (1.5)
Other language	27 (2.3)
Treatments received (n=1362), n (%)	
Surgery	549 (40.3)
External beam radiation therapy	428 (31.4)
Hormone therapy or androgen deprivation therapy	343 (25.2)
Brachytherapy	240 (17.6)
Active surveillance	210 (15.4)
Watchful waiting	150 (11)
Chemotherapy	27 (2)
Complementary and alternative therapy	31 (2.3)
High-frequency ultrasound therapy	19 (1.4)
Cryotherapy	12 (0.9)
Immune therapy	2 (0.1)
Other	89 (6.3)
I don’t know	19 (1.4)
Stage of cancer journey (n=1139), n (%)	
Follow-up monitoring after treatment	719 (63)
On active surveillance or watchful waiting	292 (26)
Recently finished treatment, but have not had any follow-up visits	47 (4)
Currently getting treatment for recurrent cancer	36 (3)
Finished treatment for recurrent cancer (less than 3 months)	24 (2)

Characteristic	Count
Receiving treatment for metastatic disease	23 (2)
Other	242 (17.2)
General health (n=1316), n (%)	
Very good	506 (38.4)
Good	727 (55.2)
Poor	77 (5.9)
Very poor	6 (0.5)

Patterns of Internet Use

Table 2 shows the patterns of internet access and use. A total of 82% of respondents were internet users. The majority of respondents (70.7%) used the internet at least once a day, had home internet access (84.3%), and accessed the internet from a desktop or laptop computer (75.7%) through broadband or wireless (47%) internet connection.

Health-Related Internet Use and Confidence

As shown in Table 2, 65% (n=784) reported that they used the internet as a source of information about prostate cancer. With respect to confidence in using information from the internet to help make decisions related to prostate cancer, 40.2% (n=521) of the respondents agreed or strongly agreed with the statement, "I feel confident in using information from the Internet to help make health decisions related to my prostate cancer," 33.5% (n=386) were undecided or disagreed, and about one-quarter (26.3%; n=341) reported that they did not use information from the internet to make decisions related to their prostate cancer. Findings from hypothesis testing provide support of the construct validity of the single item eHealth confidence variable. eHealth confidence was positively correlated with the frequency of internet use (r=0.1, *P*=.003) and negatively correlated with "not being comfortable using a computer or mobile device" (r=–0.15, *P*<.001). In addition, eHealth confidence was negatively correlated with known components of eHealth literacy when framed as barriers. These include "not knowing how to judge the quality of the information or what information to trust" (r=–0.2, *P*<.001), "not knowing how what information applied to me" (r=–0.2, *P*<.001), "not knowing how or where to search for information" (r=–0.2, *P*<.001), and "having difficulty finding information that I could understand" (r=–0.2, *P*<.001).

Table 2. Patterns of internet use (n=1362).

Characteristic	Count, n (%)
Internet use (n=1310)	1071 (81.8)
Frequency of internet use (n=1320)	
At least once a day	910 (68.9)
At least once a week (but not every day)	143 (10.8)
At least once a month (but not every week)	25 (1.9)
Less than once a month	15 (1.1)
Type of internet access (n=1362)	
A regular dial-up telephone line	61 (4.5)
Broadband such as DSL[a] or cable	640 (47.0)
A cellular network (eg, cell or smartphone)	260 (19.1)
A wireless network (Wi-Fi)	639 (46.9)
Not sure	20 (1.5)
Devices used to access internet (n=1362)	
Computer (desktop/laptop)	1031 (75.7)
Tablet	465 (34.1)
Cell phone or smart phone	411 (30.2)
Home internet access (n=1286)	1084 (84.3)
Used the internet as a source of prostate cancer information (n=1203)	784 (65.1)
Confident using information from the internet to help make decisions about prostate cancer (n=1296)	521 (40.2)

[a]DSL: digital subscriber line (the way a computer connects to the internet at high speeds using telephone lines).

Factors Correlated With Health-Related Internet Use

As shown in Table 3, factors that correlated significantly (P<.05) with health-related internet use were younger age, higher education, higher income, urban residence, broadband internet access, frequency of internet use, accessing the internet from multiple devices, confidence in using information from the internet to help make health decisions related to prostate cancer, active information seeking, active decision making role, and greater number of unmet supportive care needs.

Table 3. Factors associated with health-related internet use (total number of health-related internet users=784/1203 valid responses).

Variables (with number of valid responses for both variables) and value	Health-related internet use		*P* value
	Yes	No	
Age (years; n=1173), n (%)			<.001
43-65	316 (26.9)	79 (6.7)	
66-75	340 (29.0)	196 (16.7)	
≥76	108 (9.2)	134 (11.4)	
Education (n=1169), n (%)			<.001
Primary/secondary	176 (15.1)	180 (15.4)	
College or university	587 (50.2)	226 (19.3)	
Income (n=1091), n (%)			<.001
<CAD 40,000	136 (12.5)	137 (12.6)	
CAD 40,001-CAD 80,000	274 (25.1)	145 (13.3)	
≥CAD 80,001	307 (28.1)	92 (8.4)	
Area of residence (n=1174), n (%)			.005
Rural, town, or country	252 (21.5)	167 (14.2)	
Urban or suburban	516 (44.0)	239 (20.4)	
Broadband use (n=1203), n (%)			<.001
Yes	478 (39.7)	132 (11.0)	
No	306 (25.4)	287 (23.9)	
Internet use frequency (n=1015), n (%)			<.001
Every day	662 (65.2)	197 (19.4)	
Not every day	91 (9.0)	65 (6.4)	
Number of devices (n=962), n (%)			<.001
Computer only	284 (29.5)	137 (14.2)	
Computer+mobile device	435 (45.2)	106 (11.0)	
Electronic health confidence (n=903), n (%)			<.001
Confident	441 (48.8)	58 (6.4)	
Not confident	284 (31.5)	120 (13.3)	
Information seeking role (n=1150), n (%)			<.001
Active role	697 (60.6)	240 (20.9)	
Passive role	72 (6.3)	141 (12.3)	
Decision making role (n=1147), n (%)			.003
Active role	700 (61.0)	342 (29.8)	
Passive role	55 (4.8)	50 (4.4)	
Total unmet needs (n=1203), mean (SD)	4.62 (6.97)	3.58 (6.12)	.011

Multivariable Logistic Regression

All factors found to be significantly associated with health-related internet use in the univariable analyses were entered into the model. Table 4 shows the results of the multivariable logistic regression analysis. Our results indicate that the odds of using the internet as a health resource were higher among respondents who were active information seekers

(OR: 4.57, 95% CI 2.5-8.3), were confident using information from the internet to make health decisions (OR: 3.56, 95% CI 2.27-5.59), had broadband internet access (OR: 1.76, 95% CI 1.14-2.7), and had more unmet supportive care needs (OR: 1.05, 95%CI 1.02-1.09).

Table 4. Main effects model resulting from multivariable logistic regression.

Variables	Parameter estimate	Standard error	*P* value	Odds ratio	95% CI for odds ratio	
					Lower	Upper
Active information seekers	1.51	0.28	0.000	4.51	2.60	7.83
Confidence in using internet information to help make health decisions	1.29	0.23	0.000	3.63	2.32	5.67
Broadband internet access	0.58	0.22	0.008	1.78	1.16	2.74
Total number of unmet supportive care needs	0.05	0.02	0.004	1.05	1.02	1.09
Constant	–0.86	0.314	0.006	0.425		

Most Commonly Used Websites

A total of 22% of respondents (n=308) provided 359 valid responses to the open-ended question “Are there any specific Internet site(s) that you like to go to for information related to prostate cancer or its treatments? If Yes, please specify.” Responses that were not valid consisted of unspecified sites (eg, “health website”, search engines etc). As shown in Table 5, of the 10 categories of websites that were identified, the most commonly reported websites were national cancer agencies, government, or hospitals. The top three most commonly mentioned national cancer agencies or government websites were the Canadian Cancer Society (n=40), Prostate Cancer Canada (n=23), and the American Cancer Society (n=10). The top three most commonly reported hospital affiliated websites were Mayo Clinic (n=51) BC Cancer Agency (n=14) and Johns Hopkins (n=8). American hospital-affiliated websites accounted for over 50% of all responses in this latter category.

Table 5. Types of websites most commonly used as a source of prostate cancer information.

Order	Website type	Count, n (%)
1	National cancer agencies and government websites	133 (37.0)
2	Hospital affiliated websites	117 (32.5)
3	For-profit health information websites and news sites	39 (10.9)
4	Websites with user-generated content (eg, Youtube, Wikipedia)	15 (4.2)
5	Online support groups	13 (3.6)
6	Professional medical association websites	12 (3.3)
7	Alternative and complementary therapies	12 (3.3)
8	Academic journals, bibliographic databases	10 (2.8)
9	Personal websites	5 (1.4)
10	Electronic health record websites	3 (0.8)

Features and Information Wanted in a Website

Respondents were also asked to indicate what features and information they would want in a website for men with prostate cancer and their families. As shown in Table 6, the top three features that over 50% of respondents wanted in such a website were a library of topics (n=893, 65.6%), tools to help select treatment options (n=815, 59.8%), and tools to help navigate the prostate cancer journey (n=698, 51.2%). As shown in Table 7, the information that ≥50% of respondents wanted in such a website included information on prostate cancer treatments (n=916, 65%), what is prostate cancer and its natural progression (n=904, 64%), how to manage side effects (n=858, 61%), personally relevant information (n=769, 55%), and latest research (n=746, 53%).

Table 6. Features wanted in a website for men with prostate cancer and their families.

Order	Feature	Count, n (%)
1	A library of topics	893 (65.6)
2	Tools to help me select treatment options	815 (59.8)
3	Tools to help me navigate the prostate cancer journey	698 (51.2)
4	Links to trusted websites	649 (47.7)
5	Tools to help me assess my health status (eg, symptom assessment)	612 (44.9)
6	Tools to monitor/track changes in my health	546 (40.1)
7	Tools to record my health information (eg, PSA[a])	490 (36.0)
8	Appointment reminders (eg, by email or text)	408 (30.0)
9	Online forum to exchange info and support with other patients	389 (28.6)
10	Show your future care plan (eg, survivorship care plan)	362 (26.6)
11	Access to peer support groups	321 (23.6)
12	Tools to manage appointments (eg, calendar)	306 (22.5)
13	Tools to manage contacts	256 (18.8)

[a]PSA: prostate-specific antigen.

Table 7. Information wanted in a website for men with prostate cancer and their families.

Order	Feature	Count, n (%)
1	What prostate cancer treatments are available	916 (67.3)
2	What is prostate cancer and its natural progression	904 (66.4)
3	How to manage side effects	858 (63.0)
4	Information recommended based on my personal situation	769 (56.5)
5	Latest research	746 (54.8)
6	My chances of survival and/or cure	697 (51.2)
7	Explanations of what my doctors told me	627 (46.0)
8	How to obtain a second opinion	623 (45.7)
9	Alternative and complementary therapies	580 (42.6)
10	Wellness programs (eg, exercise, nutrition)	578 (42.4)
11	Information and access to clinical trials	523 (38.4)
12	Information to help my family deal with the prostate cancer	468 (34.4)
13	Emotional support for dealing with prostate cancer	407 (29.9)
14	Community support services in my area	361 (26.5)

Discussion

Principal Findings

This study examined the patterns of, and factors associated with, the use of the internet as a source of health information among Canadian men with prostate cancer, and the features and information wanted in a website. A total of 82% of respondents in this sample were internet users. The prevalence of internet use in this sample is consistent with the 2016 internet use rates among Canadians within this age group, which showed that 81% of people aged 65-74 were internet users [12]. Although the majority of the respondents in our sample used the internet, less than half (47%) accessed the internet through broadband and less than 30% accessed the internet with a mobile device, with the most using desktop computers to access the internet. These access patterns are also consistent with current broadband and mobile phone usage rates among Canadian seniors, which are still considerably lower than those in younger age groups [12,25].

Our findings suggest that sociodemographic factors, namely age, education, and income, may not play a significant role in determining health-related internet use among prostate cancer survivors in Canada when other factors are considered. Similarly, a previous analysis [15] of the US National Health Information Trends Survey from 2003 to 2008 showed no statistically significant differences in age, education, or

racial/ethnic aspects in health-related internet use patterns among cancer survivors, with the exception of emailing providers [15]. Survivors who were younger and had higher education were more likely to email their providers. Likewise, another analysis [11] of the 2000 Canadian National household internet survey found no association between health-related internet use and age, education, or income. Other studies suggest that there is still a gap in health-related internet use based on age and socioeconomic factors in North America [10,17]. However, these studies focused their analyses on internet users [10,17]. Chou et al [15] reported an independent association with educational attainment in their sample of cancer survivors. Similarly, our univariable analysis showed that age, education, and income were associated with health-related internet use among our sample of prostate cancer survivors.

However, our findings suggest that there is a gap in health-related internet use among prostate cancer survivors in Canada based on broadband access, which may reflect rurality [26]. Previous research has shown that urban Canadians are 1.5 times more likely to use the internet than rural Canadians [26]. In our model, respondents who had broadband access were 2.74 times more likely to use the internet as a health resource than those who did not. Broadband, as opposed to traditional dial-up, provides improved internet services, including faster browsing and downloading as well as telephone, radio, and videoconferencing. Not having access to broadband considerably limits one's quality of internet and access to essential services. In 2000, Johnson et al [11] did not find an association between high-speed internet and health-related internet use in the Canadian national household internet survey. However, longer duration of internet use was associated with decreased likelihood of using the internet as a health resource [11]. Residential broadband was launched in Canada in 1998; hence, most Canadians would have been using dial up then, which would have been costly and unstable over longer periods of use [27]. In a 2015 qualitative study, patients with prostate cancer living in remote areas of British Columbia explained that they did not like to use the internet as a health resource because it was not reliable [28]. In 2016, Canada's telecom agency declared broadband a basic service [29], along with a commitment to increase targets for download and upload speeds and extend access to regions without access. Our findings provide further evidence in favor of universal broadband internet access. Future research should investigate the impact of this policy.

In our model, the factor associated with the highest odds of using the internet as a health resource was being an active information seeker. Respondents who were information seekers were 4.6 times more likely to use the internet as a health resource than passive information seekers. A 2018 systematic review [30] of men's health-seeking behavior that largely focused on prostate cancer found that the internet was the primary source of information for active information seekers. We assessed active information seeking by asking respondents to indicate who looked for information about prostate cancer and its treatment for them. The options included (1) "I did not want any information," (2) "someone else did most or all of the looking," (3) "I did some of the looking and someone else did some of it for me," and (4) "I did most or all of the looking myself." We dichotomized the responses to create a binary variable by combining (1) and (2) to reflect passive information seekers and (3) and (4) to reflect active information seekers. This variable reflects the difference between two coping styles categorized by Miller as "Monitors" and "Blunters" [31]. Monitors focus on acquiring information to help them problem solve and reduce uncertainty and are less satisfied with the information they receive from health care providers [31]. Blunters use distraction to avoid threatening information and are typically satisfied with the amount of information they receive from their health care providers [31]. Research suggests that psychoeducational interventions for cancer patients are most effective when the level and type of information are consistent with the individual's monitoring style and the demands of the health threat [32].

The factor next most strongly correlated with health-related internet use was eHealth confidence. Respondents who were confident using information from the internet to help make health decisions related to their prostate cancer were used 3.6 times more likely to use the internet as a health resource. eHealth confidence was measured using the final item of the eHEALS eHealth literacy measure [20]. eHealth literacy is defined as the ability to seek, find, understand, and appraise health information from electronic sources and apply that knowledge to solve a health problem or make a health-related decision [20]. The final item of the eHEALs scale is intended to capture one's confidence in using information from the internet to help make health decisions. The level of agreement with this item in this sample (40.2%) is similar to that reported in a sample of older adult internet users in the United States (43.1%) [16]. Other studies have found an association between eHealth literacy, confidence, and health-related internet use. In 2003, Mead et al [33] found that positive outcome expectancy (eg, patients' belief that it would enable them to better deal with their health), previous use of health websites, and positive self-efficacy (patients' confidence in their ability to use the technology) were the strongest predictors of patient-reported interest in getting health information from the internet in the United Kingdom [33]. Ten years later, Tennant et al [16] reported that greater eHealth literacy was associated with greater use of Web 2.0 and social media for health information among a sample of Americans.

As shown in other studies of cancer patients' use of online resources [34], unmet supportive care needs were also associated with health-related internet use. For each unit increase in unmet supportive care needs, the odds ratio of using the internet as a health resource increased by a factor of 1.05. This may suggest that use of the internet by patients with prostate cancer is problem focused, driven by a need to find information to address a specific issue or problem. This finding has important implications for the design and evaluation of health information websites, as it suggests that websites should be designed to address the specific supportive care needs of patients. Respondents in our sample indicated that they want prostate cancer websites to include information on treatment options and side effects and how to manage them, as well as features that help them make treatment decisions and navigate the prostate cancer journey. This type of task-oriented internet use

that is motivated by specific needs also suggests that once prostate cancer patients' needs are met by a Web resource, they may be unlikely to keep using it, unless a new need arises. Therefore, discontinued use of a Web resource may not necessarily be a failing of the site's design but rather a potential logical reaction to changing needs and circumstances [35]. Hence, website evaluation metrics should focus on assessing whether users' specific needs have been met.

Problematically, our findings revealed that the majority of prostate cancer patients' in our sample lacked confidence in using information from the internet to make health decisions related to their prostate cancer. Research suggests that seniors have lower eHealth literacy than their younger counterparts [36]. One study found that seniors felt confident in their ability to use the internet to search for health information, but less confident in their ability to assess the quality of that information [16]. eHealth literacy interventions may help. An intervention study by Xie [37] showed that a 2-week eHealth literacy training program increased seniors knowledge, skills, and eHealth literacy efficacy [37]. At the same time, there is a need to improve the quality of prostate cancer websites. When systematically reviewed by Black and Penson in 2006 [38] and again by Kobes et al in 2018 [39], websites containing information on prostate cancer were found to be lacking in currency, attribution, balance of evidence, and comprehensiveness. In another study, only 3 of 62 websites containing information on prostate cancer treatment options were written below the recommended high school reading level [40]. In yet another study, Genova and Bender [41] found that websites containing information on prostate cancer treatment did not present information in a useful or credible way for patients. Of the 35 websites examined, the average communication quality score was 24 of 50, and less than 50% included content on risk communication, usefulness, and scientific value.

This study has certain limitations. First, as this study was cross-sectional, only association and not causation can be inferred. We aimed to obtain a representative sample of prostate cancer survivors in Canada by recruiting a random sample of patients with prostate cancer from four provincial cancer registries. It is possible that the individuals who chose not to participate in the survey had different characteristics, experiences, and attitudes regarding using the internet as a health resource. It is also possible that patterns of internet use among prostate cancer survivors have increased since then. Nonetheless, our findings are comparable to other studies involving patients with prostate cancer and current Canadian population norms. Our study was also limited by the use of the opt-in methodology used by one provincial cancer registry, which reduced the overall response rate. We encourage researchers to argue in favor of opt-out recruitment methods for future research. We also assessed only one parameter of the validity of the eHealth confidence measure. Further work is needed to establish the validity of this single item as a measure of eHealth confidence. Lastly, we did not ask for respondents' views on patient portals or personal health records, which are growing in demand. Future research should also explore the internet use trends among French-speaking Canadian patients with prostate cancer and those from diverse ethnic backgrounds.

Conclusion and Implications

The internet has the potential to serve as a health information resource for the majority of Canadian men with prostate cancer. In our sample, over two-thirds of patients with prostate cancer used the internet as a health resource. However, one-third of patients with prostate cancer in our sample did not want to use the internet as a health resource, and more than half were not confident in using the internet to make health decisions. Clinicians and educators should not assume that because of their age, education, or income, men with prostate cancer are not interested in using the internet as a health resource. Rather, they should be aware of the importance of being an active information seeker, having confidence in using information from the internet to make health decisions, and having broadband access and unmet supportive care needs as determinants of health-related internet use. These findings also show that men are looking for information on the internet about treatment options, disease progression, and management of side effects, and want websites to include features that help them make treatment decisions and navigate the prostate cancer journey. Future work should examine eHealth literacy interventions as a means to boost men's confidence in using information from the internet for prostate cancer decision making and design websites that include the information and features that patients with prostate cancer want most.

Acknowledgments

This work was awarded by Prostate Cancer Canada and is proudly funded by the Movember Foundation.

Conflicts of Interest

None declared.

References

1. International Agency for Research on Cancer. Lyon, France: International Agency for Research on Cancer; 2018. GLOBOCAN Global Cancer Statistics URL: https://gco.iarc.fr/today/fact-sheets-cancers [accessed 2019-08-12]
2. Ferlay J, Shin H, Bray F, Forman D, Mathers C, Parkin DM. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. Int J Cancer 2010 Dec 15;127(12):2893-2917 [FREE Full text] [doi: 10.1002/ijc.25516] [Medline: 21351269]

3. Canadian Cancer Statistics: A 2018 special report on cancer incidence by stage. Toronto, ON: Canadian Cancer Society; 2018. URL: http://www.cancer.ca/~/media/cancer.ca/CW/cancer%20information/cancer%20101/Canadian%20cancer%20statistics/Canadian-Cancer-Statistics-2018-EN.pdf?la=en [accessed 2019-08-12]
4. Feldman-Stewart D, Brennenstuhl S, Brundage MD, Siemens DR. Overall information needs of early-stage prostate cancer patients over a decade: highly variable and remarkably stable. Support Care Cancer 2008 Oct 2;17(4):429-435. [doi: 10.1007/s00520-008-0514-1]
5. Jefford M, Tattersall MH. Informing and involving cancer patients in their own care. The Lancet Oncology 2002 Oct;3(10):629-637. [doi: 10.1016/s1470-2045(02)00877-x]
6. Bender JL, Hohenadel J, Wong J, Katz J, Ferris LE, Shobbrook C, et al. What patients with cancer want to know about pain: a qualitative study. J Pain Symptom Manage 2008 Feb;35(2):177-187. [doi: 10.1016/j.jpainsymman.2007.03.011] [Medline: 18158232]
7. Feldman-Stewart D, Brundage MD, Hayter C, Groome P, Curtis Nickel J, Downes H, et al. What Questions Do Patients with Curable Prostate Cancer Want Answered? Med Decis Making 2000;20(1):7-19. [doi: 10.1177/0272989x0002000102]
8. Bender JL, O'Grady L, Jadad AR. Supporting cancer patients through the continuum of care: a view from the age of social networks and computer-mediated communication. Curr Oncol 2008 Aug 25;15 Suppl 2:s107.es42-s107.es47 [FREE Full text] [doi: 10.3747/co.v15i0.270] [Medline: 18769582]
9. Strecher V. Internet methods for delivering behavioral and health-related interventions (eHealth). Annu Rev Clin Psychol 2007;3:53-76. [doi: 10.1146/annurev.clinpsy.3.022806.091428] [Medline: 17716048]
10. Jacobs W, Amuta AO, Jeon KC. Health information seeking in the digital age: An analysis of health information seeking behavior among US adults. Cogent Social Sciences 2017 Mar 13;3(1):1302785. [doi: 10.1080/23311886.2017.1302785]
11. Johnson E, Johnson C S. Internet Use and Access to Health Information among Canadians: Are the Elderly on the Sidelines? J Gerontol Geriatr Res 2016;5:6. [doi: 10.4172/2167-7182.1000367]
12. The Daily. Ottawa, Ontario: Statistics Canada; 2017. Life in the fast lane: How are Canadians managing? URL: https://www150.statcan.gc.ca/n1/daily-quotidien/171114/dq171114a-eng.htm?HPA=1 [accessed 2018-07-31]
13. Feldman-Stewart D, Tong C, Brundage M, Bender J, Robinson J. Prostate cancer patients' experience and preferences for acquiring information early in their care. Can Urol Assoc J 2018 May 02;12(5):E219-E225 [FREE Full text] [doi: 10.5489/cuaj.4754] [Medline: 29405903]
14. Nagler RH, Gray SW, Romantan A, Kelly BJ, DeMichele A, Armstrong K, et al. Differences in information seeking among breast, prostate, and colorectal cancer patients: results from a population-based survey. Patient Educ Couns 2010 Dec;81 Suppl:S54-S62 [FREE Full text] [doi: 10.1016/j.pec.2010.09.010] [Medline: 20934297]
15. Chou WS, Liu B, Post S, Hesse B. Health-related Internet use among cancer survivors: data from the Health Information National Trends Survey, 2003-2008. J Cancer Surviv 2011 Sep 20;5(3):263-270. [doi: 10.1007/s11764-011-0179-5] [Medline: 21505861]
16. Tennant B, Stellefson M, Dodd V, Chaney B, Chaney D, Paige S, et al. eHealth literacy and Web 2.0 health information seeking behaviors among baby boomers and older adults. J Med Internet Res 2015 Mar 17;17(3):e70 [FREE Full text] [doi: 10.2196/jmir.3992] [Medline: 25783036]
17. Estacio EV, Whittle R, Protheroe J. The digital divide: Examining socio-demographic factors associated with health literacy, access and use of internet to seek health information. J Health Psychol 2017 Feb 01;24(12):1668-1675. [doi: 10.1177/1359105317695429]
18. Dillman D, Smyth J, Christian L. Internet, Phone, Mail, And Mixed-mode Surveys: The Tailored Design Method. New Jersey: Wiley; 2019.
19. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: Guidelines for Reporting Observational Studies. PLoS Med 2007 Oct 16;4(10):e296. [doi: 10.1371/journal.pmed.0040296]
20. Norman CD, Skinner HA. eHEALS: The eHealth Literacy Scale. J Med Internet Res 2006 Nov 14;8(4):e27. [doi: 10.2196/jmir.8.4.e27]
21. Schofield P, Gough K, Lotfi-Jam K, Aranda S. Validation of the Supportive Care Needs Survey-short form 34 with a simplified response format in men with prostate cancer. Psychooncology 2012 Oct 29;21(10):1107-1112. [doi: 10.1002/pon.2016] [Medline: 21800397]
22. Westen D, Rosenthal R. Quantifying construct validity: Two simple measures. Journal of Personality & Social Psychology 2003;84(3):608-618. [doi: 10.1037//0022-3514.84.3.608]
23. Hosmer Jr DW, Lemeshow S, Sturdivant RX. Applied Logistic Regression. New Jersey: Wiley; 2019.
24. Hsieh H, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
25. Communications MR2CCSAOFC|CRTC. 2017 Communications Monitoring Report. Ottawa, Canada: Canadian Radio-Television and Telecommunications; 2017. Canada's Communication System: An Overview for Canadians URL: https://crtc.gc.ca/eng/publications/reports/policymonitoring/2017/index.htm [accessed 2019-07-31]

26. McKeown L, Noce A, Czerny P. Rural and Small Town Canada Analysis Bulletin. Ottawa, Canada: Statistics Canada; 2007. Factors Associated with Internet Use: Does Rurality Matter? URL: https://www150.statcan.gc.ca/n1/en/pub/21-006-x/21-006-x2007003-eng.pdf?st=cGK9qVQJ [accessed 2019-07-31]
27. Lie E. Workshop on Promoting Broadband. Geneva: International Telecommunication Union; 2003. Promoting Broadband: The Case of Canada URL: https://www.itu.int/osg/spu/ni/promotebroadband/casestudies/canada.pdf [accessed 2019-07-31]
28. Flora P, Bender J, Pai H, Hodgson S, Miller A. Role of online resources in prostate cancer self-management: Perspectives of patients, caregivers and health care professionals. : Society of Behavorial Medicine; 2016 Presented at: Society of Behavioral Medicine; March 30-April 2, 2016; Washington, DC.
29. Kupfer M. CBC News. Ottawa, Canada; 2016 Dec 21. CRTC declares broadband internet access a basic service URL: https://www.cbc.ca/news/politics/crtc-internet-essential-service-1.3906664 [accessed 2019-07-31]
30. Saab MM, Reidy M, Hegarty J, O'Mahony M, Murphy M, Von Wagner C, et al. Men's information-seeking behavior regarding cancer risk and screening: A meta-narrative systematic review. Psychooncology 2018 Feb 18;27(2):410-419. [doi: 10.1002/pon.4506] [Medline: 28728212]
31. Miller SM. Monitoring versus blunting styles of coping with cancer influence the information patients want and need about their disease. Implications for cancer screening and management. Cancer 1995 Jul 15;76(2):167-177. [doi: 10.1002/1097-0142(19950715)76:2<167::aid-cncr2820760203>3.0.co;2-k]
32. Roussi P, Miller SM. Monitoring style of coping with cancer related threats: a review of the literature. J Behav Med 2014 Feb 2;37(5):931-954. [doi: 10.1007/s10865-014-9553-x]
33. Mead N, Varnam R, Rogers A, Roland M. What Predicts Patients' Interest in the Internet as a Health Resource in Primary Care in England? J Health Serv Res Policy 2016 Jun 24;8(1):33-39. [doi: 10.1177/135581960300800108]
34. Bender J, Jiang H. Online communities: Prevalence and predictors of use and reasons for non-use. London, England: Medicine 20; 2013 Presented at: Medicine 2.0; September 23-24, 2013; London, England.
35. Massimi M, Bender J, Witteman H, Ahmed O. Life transitions and online health communities. New York, USA: ACM Press; 2014 Presented at: Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing; February 15-19, 2014; Baltimore, Maryland p. 14. [doi: 10.1145/2531602.2531622]
36. Choi NG, DiNitto DM. The Digital Divide Among Low-Income Homebound Older Adults: Internet Use Patterns, eHealth Literacy, and Attitudes Toward Computer/Internet Use. J Med Internet Res 2013 May 02;15(5):e93. [doi: 10.2196/jmir.2645]
37. Xie B. Effects of an eHealth Literacy Intervention for Older Adults. J Med Internet Res 2011 Nov 03;13(4):e90. [doi: 10.2196/jmir.1880]
38. Black PC, Penson DF. Prostate Cancer on the Internet—Information or Misinformation? Journal of Urology 2006 May;175(5):1836-1842. [doi: 10.1016/s0022-5347(05)00996-1]
39. Kobes K, Harris IB, Regehr G, Tekian A, Ingledew P. Malignant websites? Analyzing the quality of prostate cancer education web resources. Can Urol Assoc J 2018 Oct 25;12(10):344-350 [FREE Full text] [doi: 10.5489/cuaj.5084] [Medline: 29989918]
40. Ellimoottil C, Polcari A, Kadlec A, Gupta G. Readability of websites containing information about prostate cancer treatment options. J Urol 2012 Dec;188(6):2171-2175. [doi: 10.1016/j.juro.2012.07.105] [Medline: 23083852]
41. Genova J, Bender J. Communication Assessment Checklist in health: Assessment and comparison of web-based health resources. Intl J User-Driven Healthcare 2016;6(2):1-20. [doi: 10.4018/ijudh.2016070101]

Abbreviations

eHEALS: eHealth Literacy Scale
HINTs: US Health Information National Trends survey
OR: odds ratio
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

Edited by G Eysenbach; submitted 02.04.19; peer-reviewed by M Saab, PA Ingledew, A Sheon; comments to author 29.05.19; revised version received 12.08.19; accepted 18.08.19; published 19.11.19.

Please cite as:
Bender JL, Feldman-Stewart D, Tong C, Lee K, Brundage M, Pai H, Robinson J, Panzarella T
Health-Related Internet Use Among Men With Prostate Cancer in Canada: Cancer Registry Survey Study
J Med Internet Res 2019;21(11):e14241
URL: http://www.jmir.org/2019/11/e14241/
doi:10.2196/14241
PMID:31742561

Original Paper

Wise Practices for Cultural Safety in Electronic Health Research and Clinical Trials With Indigenous People: Secondary Analysis of a Randomized Clinical Trial

Marion A Maar[1], PhD; Valerie Beaudin[2], CDE, RPN; Karen Yeates[3], MD; Lisa Boesch[4], BA (Hons); Peter Liu[5], MD; Kian Madjedi[6], MA, MPhil; Nancy Perkins[7], RN; Diane Hua-Stewart[7], MPH; Faith Beaudin[2,7], BSc; Mary Jo Wabano[8], MHK; Sheldon W Tobe[1,7], MScCH, MD

[1]Faculty of Medicine, Northern Ontario School of Medicine, Laurentian University, Sudbury, ON, Canada

[2]M'Chigeeng Health Centre, M'Chigeeng First Nation, ON, Canada

[3]Department of Medicine, Queen's University, Kingston, ON, Canada

[4]Department of Research, Northern Ontario School of Medicine, Sudbury, ON, Canada

[5]University of Ottawa Heart Institute, Ottawa, ON, Canada

[6]Department of Medicine, Northern Ontario School of Medicine, Sudbury, ON, Canada

[7]Department of Medicine, Sunnybrook Health Sciences Centre, Sunnybrook Research Institute, University of Toronto, Toronto, ON, Canada

[8]Naandwechige-Gamik Health Centre, Wiikwemkoong Unceded Territory, ON, Canada

Corresponding Author:
Marion A Maar, PhD
Faculty of Medicine
Northern Ontario School of Medicine
Laurentian University
935 Ramsey Lake Rd
Sudbury, ON, P3E 2C6
Canada
Phone: 1 705 662 7233
Email: mmaar@nosm.ca

Abstract

Background: There is a paucity of controlled clinical trial data based on research with Indigenous peoples. A lack of data specific to Indigenous peoples means that new therapeutic methods, such as those involving electronic health (eHealth), will be extrapolated to these groups based on research with other populations. Rigorous, ethical research can be undertaken in collaboration with Indigenous communities but requires careful attention to culturally safe research practices. Literature on how to involve Indigenous peoples in the development and evaluation of eHealth or mobile health apps that responds to the needs of Indigenous patients, providers, and communities is still scarce; however, the need for community-based participatory research to develop culturally safe technologies is emerging as an essential focus in Indigenous eHealth research. To be effective, researchers must first gain an in-depth understanding of Indigenous determinants of health, including the harmful consequences of colonialism. Second, researchers need to learn how colonialism affects the research process. The challenge then for eHealth researchers is to braid Indigenous ethical values with the requirements of good research methodologies into a culturally safe research protocol.

Objective: A recent systematic review showed that Indigenous peoples are underrepresented in randomized controlled trials (RCTs), primarily due to a lack of attention to providing space for Indigenous perspectives within the study frameworks of RCTs. Given the lack of guidelines for conducting RCTs with Indigenous communities, we conducted an analysis of our large evaluation data set collected in the Diagnosing Hypertension-Engaging Action and Management in Getting Lower Blood Pressure in Indigenous Peoples and Low- and Middle- Income Countries (DREAM-GLOBAL) trial over a period of five years. Our goal is to identify wise practices for culturally safe, collaborative eHealth and RCT research with Indigenous communities.

Methods: We thematically analyzed survey responses and qualitative interview/focus group data that we collected over five years in six culturally diverse Indigenous communities in Canada during the evaluation of the clinical trial DREAM-GLOBAL. We established themes that reflect culturally safe approaches to research and then developed wise practices for culturally safe research in pragmatic eHealth research.

Results: Based on our analysis, successful eHealth research in collaboration with Indigenous communities requires a focus on cultural safety that includes: (1) building a respectful relationship; (2) maintaining a respectful relationship; (3) good communication and support for the local team during the RCT; (4) commitment to co-designing the innovation; (5) supporting task shifting with the local team; and (6) reflecting on our mistakes and lessons learned or areas for improvement that support learning and cultural safety.

Conclusions: Based on evaluation data collected in the DREAM-GLOBAL RCT, we found that there are important cultural safety considerations in Indigenous eHealth research. Building on the perspectives of Indigenous staff and patients, we gleaned wise practices for RCTs in Indigenous communities.

Trial Registration: ClinicalTrials.gov NCT02111226; https://clinicaltrials.gov/ct2/show/NCT02111226

(J Med Internet Res 2019;21(11):e14203) doi:10.2196/14203

KEYWORDS

mobile health; process evaluation; implementation science; Indigenous peoples; health care texting; SMS; hypertension; task shifting; community-based participatory research; DREAM-GLOBAL

Introduction

Background

There is a paucity of controlled clinical trial data based on research with Indigenous peoples. As health care moves toward the era of personalized medicine, a lack of data specific to Indigenous peoples will mean that new therapeutic methods will have to be extrapolated to Indigenous patients based on work with other populations, or that new therapies, such as those involving eHealth, will not be applicable. Appropriate, ethical, and culturally safe research can be undertaken with Indigenous communities but requires attention to good research practices that improve the overall quality of the research effort.

Engaging With Indigenous Peoples in Electronic Health

The literature shows that a participatory development process involving patient and provider end users, as well as system stakeholders, is essential to optimize the uptake of electronic health (eHealth) technologies and to support sustainable innovations in health care [1,2]. Using participatory approaches, researchers should be able to explain how their results can positively impact the communities involved in the studies. Moreover, Van Gemert-Pijnen and colleagues advocate for a holistic approach to eHealth development that empathizes the importance of understanding the intervention as a whole and "takes into account the complexity of health care and the rituals and habits of patients and other stakeholders" [2].

In other words, researchers describe the need to understand the culture of the organizations that are touched and, most importantly, the culture of the people whose life is altered in some way by the technological innovation. There are many lay definitions of culture, but for the purpose of incorporating the concept of culture into eHealth research it is useful to adopt a definition from anthropology, the field where culture has been formally studied as the "mental and physical reactions and activities that characterize the individuals of a social group" [3].

Deconstructing the role of culture is key for the success of eHealth innovations, and it is particularly essential when an innovation is planned in nonmainstream cultural contexts such as Indigenous communities.

Engaging Indigenous communities in research is possible when the research question is of current interest to the communities. When researchers are aware of emerging health issues and eHealth or mobile health (mHealth) apps that have not yet caught the attention of the general population, it is much more challenging to conduct collaborative research. Literature on how to involve Indigenous peoples in the development and evaluation of eHealth/mHealth apps in order to respond to the needs of Indigenous patients, providers, and communities is still scarce; however, the need for community-based participatory research to develop culturally safe technologies is emerging as an essential focus in Indigenous eHealth research [4-6]. To our knowledge, concrete deconstructions of the factors that affect Indigenous eHealth innovations, such as: (1) the complexity of the Indigenous health system; (2) the unique historic, jurisdictional, and geographic issues; (3) effective methods of cocreation of interventions that reflect the diversity of Indigenous cultures; and (4) the Indigenous patient, provider, community, and organizational needs to facilitate uptake; have not been previously published [6]. To evaluate eHealth innovation from these highly relevant perspectives, researchers must adopt appropriate Indigenous research approaches [4,7,8].

Evaluating Electronic Health Interventions

The Randomized Controlled Trial (RCT) is widely considered to be the gold standard research methodology for evaluating whether a cause-and-effect relationship between an intervention and an outcome may exist. This traditional model of an RCT is known as the explanatory RCT and assesses outcomes under optimally controlled conditions, minimizing bias and controlling for potential confounders. One major criticism of this model of RCT is its intrinsic limitation in accommodating real-life practice conditions, such as the diversity of participants, settings, or human factors; however, these complex variables exist and interact in eHealth interventions [9,10].

Pragmatic RCTs offer a methodology and research design which prioritizes the evaluation of effectiveness (the performance of the intervention in real-world conditions) over efficacy (whether an intervention works under ideal, controlled circumstances). In a pragmatic trial design, the same measures of effectiveness

can be used as in an explanatory RCT, however, the intervention is tested in the context of everyday practice settings, which in turn requires a rigorous analysis of complex variables reflective of real-life conditions [11]. A pragmatic trial can provide data on the uptake and sustainability of technological innovations because it accommodates the study of key implementation factors in the everyday environments of clinics and communities. Process evaluations can be incorporated into a pragmatic trial, to study implementation conditions which can clarify why different outcomes exist for the same intervention at different sites.

In Indigenous communities, context includes diversity in geography, policy, and culture of patients, providers, and the organization, which all play an important role in implementation [7,12]. Analyzing these factors makes the results of pragmatic trials more informative to other Indigenous communities.

Preparing for Electronic Health Research with Indigenous Communities

If an innovation is studied as part of a clinical research trial, then rigorous data collection and analysis put considerable demands on researchers and health care workers [13]. Adding to this workload, researchers require cultural and historical training to become ready to collaborate with Indigenous communities.

First, to be effective, researchers must gain an in-depth understanding of Indigenous determinants of health in the participating population. Indigenous health is heavily impacted by the ongoing harmful consequences of colonialism; in Canada this includes the multi-generational effects of the residential school abuses [14], rampant adverse childhood experiences [15-17], loss of life styles and language [18], and the dispossession of traditional lands, resulting in food insecurity [19], collapse of traditional economies, and increase in chronic disease [20-24]. Today, health disparities in Indigenous communities continue to be reinforced through social exclusion, discrimination, and systemic racism in health care and society overall [25].

Second, researchers need to learn how colonialism affects the research process. Willie Ermine, an Indigenous ethics scholar, states that:

> *One of the festering irritants for Indigenous peoples, in their encounter with the West, is the brick wall of a deeply embedded belief and practice of Western universality [26].*

This belief manifests in a "researcher knows best" [27] attitude with research relationships that deteriorate from problematic to antagonistic:

> *Outside experts, often with little knowledge of the realities of Aboriginal community life, were commonly in a position where they controlled all aspects of the Aboriginal research projects. These experts decided which research questions warranted investigation, which methods should be used to collect data, and how the data should be interpreted and disseminated. The resulting research projects gave little consideration to the insider perspective of Aboriginal community members, existing Indigenous knowledge, the cultural competence of the research methods used, or to collaborative interpretations. Data and results were rarely accessible to community members. Knowledge transfer strategies geared to support community action on a particular problem were absent. Commonly, at the end of a project, outside experts would recommend inappropriate or unworkable solutions to community problems [28].*

For decades, government and academic research have largely failed to improve Indigenous health [28]. Research relationships have been marred by disrespect for Indigenous world views, and Indigenous communities have experienced broken trust and unethical and oppressive conduct [26,29,30]. For many Indigenous leaders, research, when conducted by outsiders, has become a dirty word [31]. The consequence of the historic practice of using research as a vehicle for hegemony manifests today as distrust towards health research even when the research need is identified by Indigenous communities. To protect communities from harm, there is a growing movement within Indigenous circles to conduct research autonomously, based on community-perceived desires and needs [29,32].

However, researchers can establish relationships if they are committed to a community-based participatory approach [8,33], building a trust-based relationship, creating ethical space for dialogue [7,26], and respecting Indigenous culture and worldviews [32,34]. Researchers need sensitivity training supported by cultural immersion to learn protocols and traditions before they can support a culturally relevant research process [35]. The challenge then for eHealth researchers is to braid Indigenous ethical values with the requirements of good research methodologies into a culturally safe research protocol.

Collaborating With Indigenous Peoples on Randomized Clinical Trials in Electronic Health

Many researchers are unable to create an ethical space for Indigenous perspectives within the study frameworks of RCTs, which in turn prevents them from building productive research relationships with Indigenous communities. A recent systematic review showed that Indigenous peoples are indeed underrepresented in RCTs [36]. The authors suggest that:

> *Rather than sidestepping Aboriginal communities, researchers should consider participatory methods for conducting RCTs with Aboriginal communities to increase the cultural relevance of these designs and to enhance the process of implementation of RCTs for optimal recruitment, engagement and retention of participants in trials, while being sensitive to the social value and cultural traditions of Aboriginal communities [36].*

We found that an explanatory RCT approach to eHealth in Indigenous community contexts was incongruent with Indigenous epistemologies. Instead, we created a pragmatic RCT protocol for DREAM-GLOBAL. The RCT protocol was shaped by our formative implementation research with Indigenous communities using the Intervention and Research Readiness Engagement and Assessment of Community Health

Care (I-RREACH) community engagement tool. We strove to allow enough flexibility to address the unique circumstances of each of the Indigenous communities from their diverse cultural perspectives [7]. This approach helped us to build the community-researcher relationship and supported discovery in our cross-cultural research.

The DREAM-GLOBAL Clinical Trial in Indigenous Communities in Canada

DREAM-GLOBAL is a complex mHealth intervention and pragmatic RCT designed to achieve improvements in blood pressure (BP) control in low resource environments using evidence of hypertension guidelines. Briefly, the objective in Canada was to evaluate the effectiveness of an innovative mHealth program using SMS text messages and electronic transfer of BP measures from patients to providers on BP control of Indigenous peoples [12]. DREAM-GLOBAL requires changes in the way services are provided that affect patients, providers, and the local health system [37].

DREAM-GLOBAL demonstrated that innovations in health services delivery, mHealth technologies, and patient engagement could be successfully implemented in collaboration with Indigenous communities in Canada [38]. In an anonymous survey designed to monitor patients' satisfaction with the intervention, 98% of the Indigenous participants (n=165/169) stated they would recommend the DREAM-GLOBAL program to a friend or relative [38]. We consider this statistic, combined with successful trial completion in all partner communities, as a preliminary indication of our culturally appropriate approach. However, given the lack of guidelines for conducting RCTs with Indigenous communities [36] it is important to critically evaluate our approach and to share our learning to support the future development of culturally safe, collaborative eHealth and RCT research within diverse Indigenous communities.

In this paper, we analyze the large qualitative data set from our pragmatic RCT collected in six culturally diverse Indigenous communities over a period of five years to identify culturally safe research practices in pragmatic RCTs.

Methods

Data Sources

The DREAM-GLOBAL RCT was implemented in six First Nations communities, representing Cree, Anishinabek (Odawa, Ojibwa, and Potawatomi) and Mi'kmaq tribes in remote northern, rural, and periurban locations in three Canadian provinces. Trial participants self-identified as Indigenous, and further mostly as First Nations with legal Indian Status.

Over a five-year period, the research team collected interview and focus group data in all phases of the research: Data collection began with community engagement work discussions that established principles for our research [7], followed by formative research to develop the intervention [39] and a process evaluation of the RCT implementation, and finally, exit interviews at the wrap up stage of the RCT [7,38]. Formal interviews and focus groups were transcribed verbatim. We kept field notes on informal discussions during site visits and the research process in general. We also conducted a participant satisfaction survey. Table 1 provides an overview of the complete data set whereas details of associated methodology were provided in previous publications.

A total of 34 interviews and 12 focus group discussions with a total of 142 participants were held in 6 communities over the period of 5 years (Table 2).

Table 1. Qualitative data analyzed in this study.

Method	Community engagement phase (pre-RCT[a])	Implementation phase (RCT ongoing)	Close out phase (post-RCT)
Data collection tool	• I-RREACH[b] Engagement Tool [7] for Health Leaders • SMS[c] focus group discussion [39] with community members (potential RCT participants)	• Process Evaluation Key Informant Interview Tool for Health Leaders • Process Evaluation Key Informant Interview Tool for Health Care Providers	• Process Evaluation Key Informant Interview Tool for Health Leaders • Process Evaluation Key Informant Interview Tool for Health Care Providers • SMS Close Out Evaluation Tool for Community RCT Participants • Patient and provider impact of cell phone use–discussion group with health care providers and community RCT participants
Summary of data collection methods	• Interviews • Focus groups • Site visit notes • Patient/provider evaluation surveys	• Provider interviews • Site visit notes	• Provider interviews • Site visit notes • Patient evaluation (satisfaction survey) • Patient/provider interviews and discussion groups

[a]RCT: randomized controlled trial.

[b]I-RREACH: Intervention and Research Readiness Engagement and Assessment of Community Health Care.

[c]SMS: short message service.

Table 2. Interviews, focus groups, evaluations, and site visit notes.

Data collected	Total number	Total participants
Focus groups	12	104
I-RREACH[a]	9	74
SMS[b]	3	30
Interviews	34	38
I-RREACH	7	7
Process evaluation	10	11
Close out process evaluation	5	5
Impact of cell phone discussion groups	12	15
Survey evaluations	233	—[c]
I-RREACH	49	—
SMS close out questionnaires	184	—
Site visit notes	13	—
Total focus group and interview participants	—	142

[a]I-RREACH: Intervention and Research Readiness Engagement and Assessment of Community Health Care.

[b]SMS: short message service.

[c]Not applicable.

Concept of Cultural Safety That Guided the Analysis

Cultural safety is a framework for understanding power differentials between health professionals and the Indigenous peoples they serve, and the negative impact of historical, social, and political power imbalances on health disparities [40]. It provides space for health care practitioners to explore the impact of these power imbalances on health while also respecting Indigenous goals of decolonization. In Indigenous health research, the privileging of Western epistemologies and methods in research over Indigenous knowledge and experiences is not only ineffective but also unethical [8,41]. Culturally safe research requires researchers to reflect on their tacitly held values, beliefs, and practices. Researchers and the organizations that support them must learn to identify, understand, and change routine practices, habits, or behaviors that create unsafe experiences for communities and participants [42]. Most importantly, it is the Indigenous research participants (not the researchers) who determine if a research project has been culturally safe and this must be reflected in the methodology [43]. In our analysis we therefore sought out narratives that illustrated cultural safety (or lack thereof) from the perspective of Indigenous research participants.

Data Analysis

We utilized an inductive approach to the thematic analysis of the transcripts, with two researchers coding and categorizing the data for themes related to cultural safety, as expressed by participants. We then contextualized the emerging cultural safety themes around implementation issues, including the impact of the technology and delegation of tasks such as measuring patients' blood pressure to nonregulated providers (also known as task shifting) [44]. We focused on understanding what factors led to a perceived fit (or lack of fit) with cultural safety based on the perspectives of Indigenous patients and staff. Our results were also informed by field notes which documented our learnings on cultural safety throughout the trial and were based on informal conversations at the community level. Once a preliminary analysis was completed, community-based Indigenous co-researchers provided feedback on the thematic analysis during several meetings and verified the final analysis through rigorous member checking.

Ethics

Ethics review was completed by community-based First Nations REBs and university-based REBs. Community-based ethics review in First Nations communities included The Cree Board of Health and Social Services of James Bay, and the Manitoulin Anishinaabek Research Review Committee [34]. The study was formally approved by First Nations leadership through Band Council Resolutions. Academic ethics approvals include: (1) Queen's University Health Sciences and Affiliated Teaching Hospitals Research Ethics Board (DMED-1603-13); and (2) Sunnybrook Health Sciences Centre Research Ethics Board (#182-2013).

Trial Status

The RCT was completed December 2017 and outcomes are published [37].

Results

Summary

Based on our analysis, successful eHealth research in collaboration with Indigenous communities requires a focus on cultural safety that includes: (1) building a respectful relationship; (2) maintaining a respectful relationship; (3) good communication and support for the local team during the RCT;

(4) commitment to co-designing the innovation; (5) supporting task shifting with the local team; and (6) reflecting on our mistakes and lessons learned or areas for improvement that support learning and cultural safety.

Building a Respectful Relationship

Overview

Community staff linked cultural safety with the researchers' commitments to building a respectful relationship based on open and transparent in-person meetings in the community. The focus on the community's perspective, including each community's unique cultural protocols, must be established during the community engagement and implementation process.

Face to Face Always [Has] the Best Results (Community B)

Staff emphasized the importance of meeting in person with the research team, including the principal investigator (PI).

> *The commitment to come here, I think that's big.* [Community E]

Despite being a clinician in a busy hospital, taking time to visit each community on multiple occasions allowed the PI (ST) to learn directly from health staff, address concerns, and to provide background information about the research. The research coordinator (NP) was also a key participant in visits. In-person conversations established the rapport with the local team to facilitate trouble shooting of technology, recruitment, and clinical issues throughout the study. Other members of the team attended at the community level when possible.

Focus Was on Our Problems (Community B)

Cultural safety from the perspective of the participants involved focusing on community perceived issues as opposed to "researcher-knows-best" issues. For this, researchers required a basic knowledge of Indigenous determinants of health and the legacy of colonial policies [25]. Participants felt validated as researchers acknowledged colonial effects on Indigenous health by including a "discussion of multigenerational impacts" (Community B) on hypertension. Specific determinants of health that affected communities to different degrees, such as poverty, meant, "not everyone had cell phones" (Community B). The research team responded to this by providing simple phones to those in need so that everyone had the ability to participate in the project, which was important to community-based advisors.

Openness of the Researchers (Community D)

Taking the time for open dialogue and taking community perspectives seriously was one of the most often cited examples of cultural safety in DREAM-GLOBAL. Community staff appreciated the "comfortable environment to discuss the project openly and honestly" (Community D). They also stated that it was important to them that during meetings the "the atmosphere was good and friendly" (Community A), "and [community staff] felt at ease" (Community A) with the research team, who were "easy to talk to" (Community A). This encouraged local "people to share freely" (Community B) and "gain an understanding of the program and realizing there are many supports in place" (Community A).

The importance of addressing power imbalances was raised as participants stated their appreciation that the researchers were "non-judgmental…gentle and well-versed" (Community C) in working appropriately with Indigenous people. The community staff questions were taken seriously and "the research group went into detail" (Community A) about all aspects of the project, including concerns about confidentiality of health records, which were thoroughly addressed by the PI. This elicited the remark: "[you] made [it] very clear about confidentiality–Miigwetch (translation: thank you)" (Community B).

Cultural and Traditional Approach Was Taken Into Consideration (Community B)

Cultural safety is also about respecting and accepting the culture of the local clinics:

> *The team seems used to working with Indigenous communities. They accept our pace. They aren't trying to turn this into a clinic in the South.* [Community F]

Staff liked that the "people leading were positive" (Community C), "their tone was relaxing and comforting" (Community C) and they "did not feel rushed" (Community C) in their work with the community. Cultural safety also came into play at the patient level, when the team realized that the SMS messages needed to be adapted to include traditional activities and foods based on focus group feedback [39]. From the researcher perspective, this was accomplished by asking community members to co-develop the text messages:

> *We got some great ideas for incorporating traditional foods and activities in an appropriate manner [into the intervention during our community visit]. For example, something like "Keep wild meat/foods healthy by boiling, broiling, stewing and limiting added fats…"* [Field notes on community discussion during formative research in Community A]

Maintaining a Respectful Relationship

Their Visits to the Community Keeps Us Motivated (Community E)

Maintaining the relationship between the community and the researchers was an ongoing process and could not be reduced to a kick-off event. Participants explained that cultural safety in research required a focus on strengthening the collaborative relationship:

> *In First Nation communities, that's huge, relationship building. The level of comfort is there [with the research team]. People come into the room and sit and talk when the DG team is here, so that means they feel comfortable. And then, when they have that level of comfort, they're open to what you have to say.* [Community E]

Nurturing the relationship required that researchers were available to meet with all kinds of community members, leaders, and groups, not merely the gatekeepers.

> *I think it was really respectful. You met with the Elders, the band office, the CHRs (Community Health Representatives), so that was – yes, I think it just*

respects the basic collaboration criteria you have to do in…research with First Nations. [Community F]

There's a Feast for DREAM-GLOBAL (Community F)

The Indigenous cultural value of sharing is at the root of events such as community feasts. With support from the local staff, the DREAM-GLOBAL team hosted community feasts to share with the community as a whole, while also creating awareness about the project.

> *I think the feast is a good way also. I heard the community talking about "Oh yeah there's a feast for DREAM-GLOBAL." So there was some person there […] and she never heard before of the DREAM-GLOBAL study and she said 'Oh that's great, I also have high blood pressure.' I'm happy to hear what you did, like the small teaching in front. […] I think it's a [good] way to communicate.* [Community F]

Other communities also found feasts to be a culturally appropriate way to promote DREAM-GLOBAL.

> *Most people now know about it, especially since the Elders' meal yesterday, there was quite a buzz yesterday, and now a lot of people are coming to check what's going on.* [Community E]

You Tasted Our Traditional Food (Community F)

Participants emphasized the importance of positive interactions and that the openness of the researchers for engaging with the local culture and traditions also contributed to culturally safe research.

> *So the respect and wanting to learn about the culture was there I believe. […] In my opinion, certainly was there because I know you tasted our traditional food.* [Community F]

Good Communication and Support for the Local Team During the Randomized Controlled Trial

The Relationship Was Very Open (Community F)

Community staff consistently provided feedback that the key to culturally safe research was the supportive communication with the DREAM-GLOBAL team, which helped to strengthen the relationship.

> *I think the relationship was very open. […] Any issues, you always address them. You're always open to respond…* [Community F]
>
> *Any questions that arise, …[the researchers] are always available to answer our questions and provide us with assistance.* [Community C]

Timely access to the nurse research coordinator was considered an important aspect of DREAM-GLOBAL, especially the supportive troubleshooting.

> *I find the team is very supportive. If I need anything, I can quickly do a text message. I can get the information right away. I don't have to wait.* [Community D]
>
> *We always feel that you are there to support us. There's never been issues where we don't receive the feedback that we need. It's been excellent.* [Community E]

The tone of the interaction was also frequently mentioned:

> *Communications with the DREAM-GLOBAL team was all very pleasant and I feel very supported by them. […] They are approachable and it's good to have good support. It's important to have people who are understanding, friendly and encouraging.* [Community C]

Respect, openness, kindness, and support were clearly key qualities that the community staff valued in the communication.

Commitment to Co-Designing the Innovation

Balancing Culture and Trial Design

The main challenge in co-designing a culturally safe intervention was finding ways to balance the clinical trial requirements with community culture. For example, early on we were told that a strict randomized trial where some participants received treatment and others placebo was not in line with Indigenous cultural values of respect and sharing. We therefore used "active versus passive" text messages, which resulted in a reduced-treatment arm as opposed to a no-treatment arm. However, it did not stop there, as the messages were not culturally safe.

They Want Us to Rephrase all SMS Messages (Field Notes)

Our evidence-based hypertension treatment text messages underwent rigorous testing for cultural safety in each of the participating Indigenous communities. Phrasing was important for cultural safety, as Indigenous patients interpret text messages within the context of their personal experience of oppression and racism in medical institutions. Indigenous community members taught the research team to avoid message content that could be perceived as paternalistic, fear-inducing, oppressive or authoritarian.

> *[Participants] want us to rephrase all messages that "compel". So for example: "keep taking your meds as instructed…." This phrasing elicited a really emotional response and active resistance in our participants. "Don't tell me what to do like I am a kid – offer us choices and reminders instead". So these will need to be changed to "It is a good idea to take medications as indicated by your health care provider" or 'Have you taken your meds today?' etc.* [Field notes in Community A]

Many involved in the testing had strong dislikes for messages that would be acceptable to most Canadians. Cocreating a culturally safe version required unplanned formative research [39] but ensured that the messages were perceived as welcome and trustworthy. At close out, community participants reflected on the cultural appropriateness of the cocreated messages:

> *And they still, to this day […they] tell us, you know, "I've learned a lot from those messages that I wasn't aware of before."* [Community F]

> *I felt like somebody cared about my health. The messages were good reminders and were motivating.* [Community F–SMS participant feedback questionnaire]

Additional details, such as how many text messages patients were to receive and at what time of the day, were determined through dialogue with community staff. They identified an optimal frequency and timing, which was later validated as a good fit by participating patients:

> *Every couple of days when they get their text, it says something about the way they should be eating, lowering their salt intake. And they're able to work on it for that week. It's kind of like a reminder for them to keep on, you know, taking little baby steps. It's a constant little gentle reminder, nothing too harsh but at the same time it turns into a huge positive outcome for them later on.* [Community D]

Supporting Task Shifting With the Local Team

Feeling Very Closely Involved (Community F)

Staff felt it was important that the research team listened equally, and in a nonhierarchical manner, to incorporate input from all community team members:

> *There is a real partnership between the research team […] and the CHRs. They are very happy that [the PI] actually came to the community at the implementation of the study. The team regularly asks for the CHRs input etc. […] The CHRs love that they are feeling very closely involved in the project.* [Community F]

This emphasis on collaboration with all team members, including the nonregulated health care workers, was crucial when community health representatives took on new roles. Blood pressure monitoring or management was a task that shifted from nursing staff to community health representatives, which required culturally safe training and support to empower community health representatives to perform this new task:

> *At the beginning… we didn't feel comfortable because I've never done it as a CHR, but as we progressed… in the study, I really looked forward to meeting once a month, the client, the participants…CHRs never had done the blood pressure check before so I felt like one of the health professionals. I felt proud of myself to do it.* [Community F]

> *…a personal success is becoming more comfortable with teaching and the physiology of BP and how it works. And I've become more confident the more I do it.* [Community C]

The RCT proved to be positive for personal and professional development of staff and it added to their recognition as they learned new capacities and current limits.

> *And at a staff level, we see people gaining knowledge and building capacity in how to manage hypertension and chronic diseases.* [Community E-HD]

The successful support for task shifting in turn supported recruitment for the study,

> *…one of the successes was it was easy to get people to buy into the program. We didn't have to do too much in the way of PR, it was more word of mouth. People were really into leading a healthier lifestyle, it wasn't hard for them to be swayed or we didn't have to do a lot of teachings, I guess...[DREAM-GLOBAL] had a unique, innovative way of thinking about research studies, so that was one of the pluses about it. And our community was ready for it.* [Community D]

> *Within our primary care staff, it is good and we update each other every week on what we've done. And I always give an update on DREAM-GLOBAL to the staff, and that's why I get a lot of support and referrals.* [Community C]

Some communities had slow recruitment periods, but the staff had become invested and persevered:

> *Initially it was challenging, then it seems all of a sudden there was a lot of interest. […] It's all about timing. The right time, the right place. […] We saw at the beginning, we were able to get quite a few, then it was quiet, then we had an opportunity again to get more people enrolled.* [Community E]

Many felt that the expansion of the community health representative role was sustainable.

> *So for me the biggest success is to show what the CHR can do within their power and I think that's going to leave traces for the next years that we'll say well we know that you can do it and I'm sure they want to do it as well.* [Community F]

Reflecting on Our Mistakes

Need for Reflexivity

Cultural safety in research required reflexivity about the effects of the project on participating communities, staff, and patients. Reflecting on DREAM-GLOBAL, our focus on cultural safety meant that many things went well but also that we learned from our mistakes.

In for a Week, Out for a Week, You Know? (Community F)

We attempted to adhere to community collaboration, but the health system is different in each First Nation community and understanding the mix of provincial and federal services and permanent, contract, and visiting staff can be challenging for visiting researchers. In one community, we had failed to notice that the functioning of the local team was fractured by staff turnover, and lack of integration of contract and community staff. Consequently, we missed the opportunity to consult with key members in the community at the beginning, which stymied the integration of the eHealth intervention within the local health

services. Later we realized that we had to orient many more health care providers, including those who were in the community only intermittently.

> *I would have wanted the doctors to be more involved but…they couldn't really be part of it because they're always in and out of the community because they have other communities to [be] responsible [for]….CHRs have to keep reminding the MDs to refer and still they don't do so enough.* [Community F]

> *One of the major problems was that management was not included in the community discussion as to whether to accept to participate in the DG study or not. [The director] feels like the doctors went over her head.* [Community F]

Our failure to understand who could speak on behalf of the community and who could decide to integrate the eHealth innovation into the services also lead to an overestimation of hypertension as a health priority.

> *I think people just forgot…blood pressure is something not in the first priorities. Here, I think it's more like diabetes.* [Community F]

I Felt a Little Bit Out of the Loop (Community F)

The effect of our oversight snowballed and affected important aspects of the implementation of DREAM-GLOBAL, including recruitment and communication:

> *…doctors… are coming in and out of this health care system all the time. So were we catching all the [patients] who would be open to take part in the project? Probably not… Maybe not everybody knew to ask, you know?* [Community F]

> *she was told by one of the MDs that the nurses would have nothing else to do but to refer the patients to the CHRs, but… the CHRs would ask her if she had discussed the study with the patient prior to the referral; there seemed to be a misunderstanding in each person's role.* [Community F]

The shifting of new tasks to the community health representative was also initially much less accepted at this site as we were struggling with a culturally safe approach.

> *I get the feeling that when we refer to the CHR, some of [the patients] may feel like we've just passed them off to somebody else. […] So a lot of times they don't want to go see the CHR and sometimes maybe it's because they have like personal issues with some of the CHRs, I don't know. It's a small community, you know, family issues between families.* [Community F]

However, on the positive side, using our own methods of reflection and face to face visits to resolve our rocky start, we achieved a satisfactory level of implementation in the end, especially after spending more time with the community health representatives.

> *At the beginning it was slow but once we got the awareness and when we did that blood screening at the commercial center […] then the people were just coming, like flocking to our screening.* [Community F]

Discussion

Primary Findings

At the onset of the RCT, our team was committed to cultural safety in our research process and seeking the perspective of the community in all phases of the project. We often succeeded with our approach, but despite good intentions we sometimes fell short. Carving out time for reflection on these miscalculations with community members and as a research team was important to transform these experiences into learning opportunities and to resolve misunderstandings.

Listening to Indigenous community staff, elders, and patients during our community visits, we discovered that in order to prepare for clinical trial research the most significant undertaking is tending to the relationship between the researchers and the community at various stages of the work (Figure 1).

The relationship building starts with an engagement process, where researchers and communities learn about topics such as community issues and cultural and research protocols and come to a consensus about expectations for the trial phase. Maintaining research relationships after the trial may include cocreating a plan for community presentations, report writing, advocacy for sustainability, and planning for future collaborations. In this case the relationship becomes circular, as demonstrated in Figure 2. These findings are conceptually in line with established ethical practices in Indigenous research [8,32,45,46], as well as supported by our previous research [7,27,39,44].

The findings in this study are tailored to pragmatic clinical trial research and to eHealth research with Indigenous communities, and they illustrated, with selected narratives, that researchers can gain a better understanding of various Indigenous perspectives related to RCT practices and methodologies. The narratives underscore the important cultural value of relationship building shared by all participating Indigenous groups, and the role that researchers' commitment, community immersion, good communication, and supportive attitude plays in strengthening their relationship with the community. Electronic Health requires a thorough assessment of how the intervention will affect local workflow in a way that is empowering to the community, with task shifting one important way to achieve this goal. Finally, time for critical reflection with the community representatives and within the research team to understand good practices and mistakes are important aspects of culturally safe research.

Figure 1. Research relationship stages.

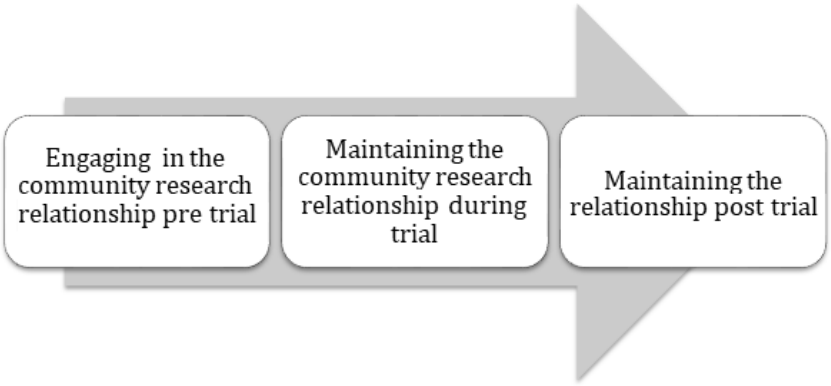
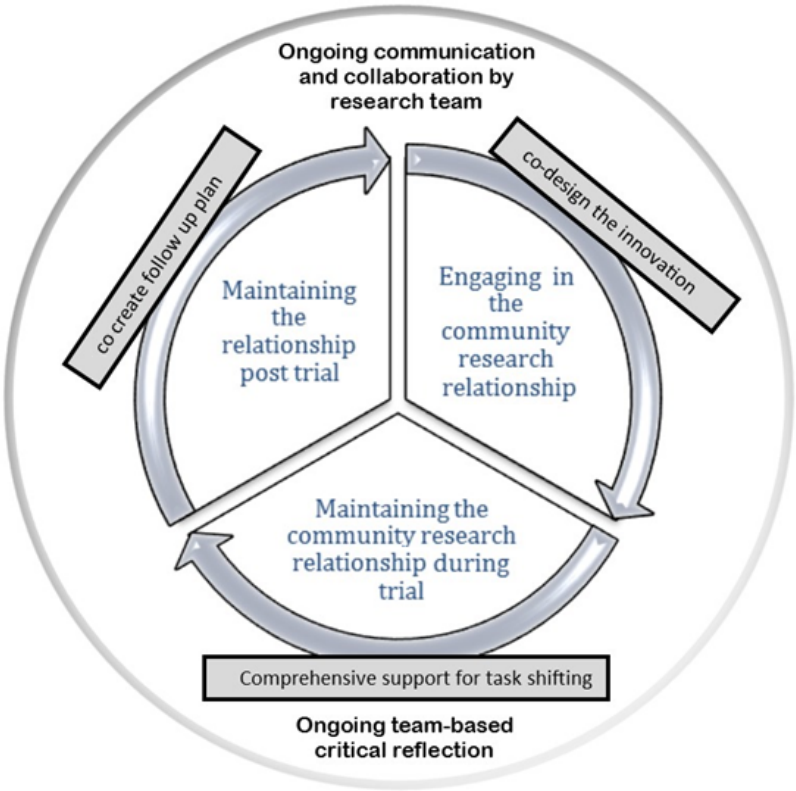


Figure 2. Maintaining research relationships.



Building on the literature and the results of our research with six culturally diverse Indigenous communities, we formulated wise practices for cultural safety in RCT and eHealth research. We do not claim that these are exhaustive practices, nor that they apply exactly as stated to all Indigenous communities. Instead, we invite researchers and Indigenous community partners who collaborate on RCT and eHealth research to review these practices as a starting point. We believe most points will resonate, however, some may need to be collaboratively added or modified to the unique characteristics of each Indigenous community and the corresponding research project.

Wise Practices for Culturally Safe Randomized Clinical Trial Research

Focus on Researcher Readiness

Researchers need to learn about Indigenous issues and take this up as part of life-long learning. This process may include completing training (including online) on cultural safety, self-reflection, ethical space dialogue, Indigenous culture, history, and treaties to gain capacity in this area [7,35]. Course work should be complemented with experiential work and immersion in Indigenous culture and communities whenever possible.

Changing the research lexicon, from investigator to researcher and from trial to study may also be helpful. These words are less charged with meanings that can lead to reduced trust. Learning about the specific communities involved is also important.

Reflective Research Practice

Reflection is a key component of culturally safe care, and researchers should participate in regular ongoing reflection on cultural safety in research [40]. Similarly, critical reflection should happen on a personal level, as a research team, and as a

project team together with the local Indigenous team. Open dialogue should support this activity and translate into adjustments towards more culturally safe approaches as needed. It may be helpful to reflect in pairs, so that if the meaning is not fully understood dialogue can lead to increased understanding.

Principal Investigators and Study Coordinators Need to Build a Relationship With Communities

Both the principal investigators and study coordinators should spend significant time in face-to-face meetings in the community during all phases of the project to build a respectful rapport and answer all questions openly and honestly in an informal manner. This will ensure that both the key decision makers and support personnel are aware of and understand community wishes, concerns, and opportunities related to the project. Meetings should include local clinical champions who should be well integrated in the community [7].

Trial Should Include Some Benefits for all Participants

A treatment versus no treatment control design is often deemed to be culturally inappropriate or unethical from the perspective of Indigenous world views. Trial design should allow for benefits to the control group by providing, for example, reduced treatment in a pragmatic trial, or delayed treatment (step wedge design trial). Small gifts to acknowledge appreciation for participation are welcome (and often a cultural practice), particularly practical ones such as healthy food boxes or gift cards for local produce retailers [27].

All Phases of the Electronic Health Intervention Are Co-Designed

The health issue, messages, graphics, research process, and all aspects of the technology must be acceptable to participants. This requires formative research to co-design the intervention in collaboration with the Indigenous team members, to ensure that potential adaptations during implementation and follow up after the trial are completed so that there are lasting benefits to the community [39,44]. While there is an understanding that there are health disparities, true co-design will likely focus on incorporating a strengths-based approach to the health issue and will build on Indigenous culture, community, and resilience.

Technology and Research Support

Technology support will be required to troubleshoot, whereas research support is needed for questions related to recruitment, blinding, etc. One key contact person who has established a relationship with the community is needed to respond quickly to community staff's questions. This person may not have all the answers but should be well connected with all members of the research team so they can get the answers in a timely fashion [38].

Support for Task Shifting

The task shifting should support local goals for self-sufficiency in community health and community empowerment. Changes in roles require that researchers learn to understand the local work dynamics and advocate for acceptable shifting of tasks within the local health system. To achieve this, training and supporting the community staff who carry out the new task, as well as the managers, is necessary.

Research Budgets Must Reflect the Nature of Community-Based Participatory Research

Budgets need to cover researcher travel to communities, community-based collaborator travel to urban meetings or conferences, and culturally appropriate local hospitality, honoraria, or gifts. It must also cover post study knowledge transfer to ensure that community members feel that they have benefitted as much as possible from the new learnings from the project. Timelines for spending funds requires flexibility to be respectful of local priorities and competing commitments to avoid overburdening the community workers.

Limitations

Our work was limited to six First Nations communities, thus there are limitations related to the generalizability of the work. However, as many of the themes were found in these culturally diverse communities that included Cree, Mi'kmaq, Pottawatomi, Ojibwa and Odawa tribes, it is likely that the criteria for cultural safety in research will strongly resonate with many other Indigenous peoples in Canada and in other countries.

Conclusions

Based on evaluation data collected over the five years of the DREAM-GLOBAL RCT, we found that there were important cultural safety considerations in Indigenous eHealth research. Building on the perspectives of Indigenous staff and patients, we gleaned wise practices for RCTs and eHealth research in Indigenous communities.

Cultural safety in eHealth research is dependent on: the quality of community engagement and collaboration on all phases of the research; sustained relationship building; respectful communication; timely implementation support; a commitment to co-design the innovation in collaboration with Indigenous partners, and support for culturally appropriate task-shifting. Finally, reflecting and learning from mistakes is needed to ensure cultural safety.

Acknowledgments

The authors would like to thank health care staff and leadership of participating First Nations communities and organizations for their collaboration. These include Aundeck Omni Kaning First Nation, Cree Nation of Chisasibi, Eel River Bar First Nation, M'Chigeeng First Nation, Noojmowin Teg Health Centre, Sheshegwaning First Nation, Wiikwemikoong Unceded Territory. The project was funded by The Canadian Institutes of Health Research, Grand Challenges Canada, and the Global Alliance for Chronic Diseases.

Conflicts of Interest

None declared.

References

1. Eysenbach G. Medicine 2.0: social networking, collaboration, participation, apomediation, and openness. J Med Internet Res 2008 Aug 25;10(3):e22 [FREE Full text] [doi: 10.2196/jmir.1030] [Medline: 18725354]
2. van Gemert-Pijnen JEWC, Nijland N, van Limburg M, Ossebaard HC, Kelders SM, Eysenbach G, et al. A holistic framework to improve the uptake and impact of eHealth technologies. J Med Internet Res 2011 Dec 05;13(4):e111 [FREE Full text] [doi: 10.2196/jmir.1672] [Medline: 22155738]
3. Boas F. Race, Language and Culture. New York: The Macmillan Company; Oct 1941.
4. Maar MA, Seymour A, Sanderson B, Boesch L. Reaching agreement for an Aboriginal e-health research agenda: the Aboriginal Telehealth Knowledge Circle consensus method. Rural Remote Health 2010;10(1):1299 [FREE Full text] [Medline: 20108996]
5. Davies J, Bukulatjpi S, Sharma S, Caldwell L, Johnston V, Davis JS. Development of a Culturally Appropriate Bilingual Electronic App About Hepatitis B for Indigenous Australians: Towards Shared Understandings. JMIR Res Protoc 2015 Jun 10;4(2):e70 [FREE Full text] [doi: 10.2196/resprot.4216] [Medline: 26063474]
6. Jones L, Jacklin K, O'Connell ME. Development and Use of Health-Related Technologies in Indigenous Communities: Critical Review. J Med Internet Res 2017 Jul 20;19(7):e256 [FREE Full text] [doi: 10.2196/jmir.7520] [Medline: 28729237]
7. Maar M, Yeates K, Barron M, Hua D, Liu P, Moy Lum-Kwong M, et al. I-RREACH: an engagement and assessment tool for improving implementation readiness of researchers, organizations and communities in complex interventions. Implement Sci 2015 May 04;10:64 [FREE Full text] [doi: 10.1186/s13012-015-0257-6] [Medline: 25935849]
8. Government of Canada. TCPS 2 (2018) - Chapter 9: Research Involving the First Nations, Inuit and Métis Peoples of Canada. 2018. URL: http://www.pre.ethics.gc.ca/eng/tcps2-eptc2_2018_chapter9-chapitre9.html [accessed 2019-03-01]
9. Macpherson H. Pragmatic clinical trials. Complement Ther Med 2004;12(2-3):136-140. [doi: 10.1016/j.ctim.2004.07.043] [Medline: 15561524]
10. Hotopf M. The pragmatic randomised controlled trial. Adv. psychiatr. treat 2018 Jan 02;8(5):326-333 [FREE Full text] [doi: 10.1192/apt.8.5.326]
11. Patsopoulos NA. A pragmatic view on pragmatic trials. Dialogues Clin Neurosci 2011;13(2):217-224 [FREE Full text] [Medline: 21842619]
12. Yeates K, Campbell N, Maar MA, Perkins N, Liu P, Sleeth J, et al. The Effectiveness of Text Messaging for Detection and Management of Hypertension in Indigenous People in Canada: Protocol for a Randomized Controlled Trial. JMIR Res Protoc 2017 Dec 19;6(12):e244 [FREE Full text] [doi: 10.2196/resprot.7139] [Medline: 29258978]
13. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
14. Truth and Reconciliation Commission of Canada. Honouring the Truth, Reconciling for the Future: Summary Report of the Final Report of the Truth and Reconciliation Commission of Canada. 2015. URL: http://publications.gc.ca/site/eng/9.800288/publication.html [accessed 2019-03-01] [WebCite Cache ID 76YGtt6AH]
15. Felitti VJ, Anda RF, Nordenberg D, Williamson DF, Spitz AM, Edwards V, et al. Relationship of childhood abuse and household dysfunction to many of the leading causes of death in adults. The Adverse Childhood Experiences (ACE) Study. Am J Prev Med 1998 May;14(4):245-258. [Medline: 9635069]
16. Bombay A, Matheson K, Anisman H. Intergenerational Trauma: Convergence of Multiple Processes among First Nations peoples in Canada. 2009. URL: https://pdfs.semanticscholar.org/09dc/806e7896d63c5f523e55f3bb3dff1709230e.pdf?_ga=2.103439494.749505452.1551457384-1993338492.1551457384 [accessed 2019-03-01] [WebCite Cache ID 76YHTRAOJ]
17. Koss MP, Yuan NP, Dightman D, Prince RJ, Polacca M, Sanderson B, et al. Adverse childhood exposures and alcohol dependence among seven Native American tribes. Am J Prev Med 2003 Oct;25(3):238-244. [Medline: 14507531]
18. Manitowabi D, Maar M. We stopped sharing when we became civilized: A Model of Colonialism as a Determinant of Indigenous Health in Canada. J. Indig. Soc. Dev 2018;7(1):1-19 [FREE Full text]
19. Settee P, Shukla S, editors. Indigenous Food Systems: Concepts, Cases, and Conversations. Toronto: Canadian Scholars; 2020.
20. Gracey M, King M. Indigenous health part 1: determinants and disease patterns. Lancet 2009 Jul 04;374(9683):65-75. [doi: 10.1016/S0140-6736(09)60914-4] [Medline: 19577695]
21. Anand SS, Yusuf S, Jacobs R, Davis AD, Yi Q, Gerstein H, et al. Risk factors, atherosclerosis, and cardiovascular disease among Aboriginal people in Canada: the Study of Health Assessment and Risk Evaluation in Aboriginal Peoples (SHARE-AP). Lancet 2001 Oct 06;358(9288):1147-1153. [doi: 10.1016/s0140-6736(01)06255-9] [Medline: 11597669]
22. Waldram JB, Herring DA, Young TK. Aboriginal Health in Canada: Historical, Cultural, and Epidemiological Perspectives (2nd Edition). Toronto: University of Toronto Press; 2006.

23. Aguiar W, Halseth R. Aboriginal Peoples and Historic Trauma: The processes of intergenerational transmission.: National Collaborating Centre for Aboriginal Health; 2015. URL: https://www.ccnsa-nccah.ca/docs/context/RPT-HistoricTrauma-IntergenTransmission-Aguiar-Halseth-EN.pdf [accessed 2019-03-01] [WebCite Cache ID 76YIKsHWG]
24. Manitowabi D, Maar M. Coping with Colonization: Aboriginal Diabetes on Manitoulin Island. In: Fear-Segal J, Tillett R, editors. Indigenous Bodies: Reviewing, Relocating, Reclaiming. New York: SUNY Press; 2013:145-159.
25. Reading C, Wien F. Health Inequalities and Social Determinants of Aboriginal Peoples' Health. Prince George: National Collaborating Centre for Aboriginal Health; 2009. URL: https://www.ccnsa-nccah.ca/docs/determinants/RPT-HealthInequalities-Reading-Wien-EN.pdf [accessed 2019-03-01] [WebCite Cache ID 76YJ3qVSR]
26. Ermine W. The Ethical Space of Engagement. Indigenous Law Journal 2007;6(1):193-203 [FREE Full text]
27. Maar MA, Lightfoot NE, Sutherland ME, Strasser RP, Wilson KJ, Lidstone-Jones CM, et al. Thinking outside the box: Aboriginal people's suggestions for conducting health studies with Aboriginal communities. Public Health 2011 Nov;125(11):747-753. [doi: 10.1016/j.puhe.2011.08.006] [Medline: 22093920]
28. Maar M, Sutherland M, McGregor L. A Regional Model for Ethical Engagement: The First Nations Research Ethics Committee on Manitoulin Island. 2007. URL: https://ir.lib.uwo.ca/aprci/112/ [accessed 2019-03-01] [WebCite Cache ID 76YKG4de1]
29. Chilisa B. Indigenous Research Methodologies. Thousand Oaks: SAGE; 2011.
30. Ellison C. Indigenous Knowledge and Knowledge Synthesis, Translation and Exchange (KSTE). 2014. URL: https://www.ccnsa-nccah.ca/docs/context/RPT-IndigenousKnowledgeKSTE-Ellison-EN.pdf [accessed 2019-03-01]
31. Smith LT. Decolonizing Methodologies: Research and Indigenous Peoples. London: Zed Books; 2012.
32. McGregor D, Restoule JP, Johnston R. Indigenous research: theories, practices, and relationships. Toronto: Canadian Scholars Press; 2018.
33. Government of Canada, Canadian Institutes of Health Research. Defining Indigenous Health Research. 2017. URL: http://www.cihr-irsc.gc.ca/e/50340.html [accessed 2019-03-01] [WebCite Cache ID 76YOE21QW]
34. Manitoulin Anishinaabek Research Review Committee. Guidelines for Ethical Aboriginal Research. Little Current: Noojmowin Teg Health Centre; 2003. URL: https://www.noojmowin-teg.ca/images/GEAR_-_FINAL.pdf [accessed 2019-03-01] [WebCite Cache ID 76YQVRujm]
35. Ermine W, Sinclair R, Bonnie J, Indigenous Peoples' Health Research Centre. The Ethics of Research Involving Indigenous Peoples. Saskatoon: Indigenous Peoples' Health Research Centre; 2004. URL: http://iphrc.ca/pub/documents/ethics_review_iphrc.pdf [accessed 2019-03-01] [WebCite Cache ID 76YQkdfPj]
36. Saini M, Quinn A. A Systematic Review of Randomized Controlled Trials of Health Related Issues within an Aboriginal Context. Prince George: National Collaborating Centre for Aboriginal Health; 2013. URL: https://www.ccnsa-nccah.ca/docs/context/RPT-ReviewRCTs-Saini-Quinn-EN.pdf [accessed 2019-03-04] [WebCite Cache ID 76clGkog6]
37. Tobe SW, Yeates K, Campbell NRC, Maar MA, Perkins N, Liu PP, et al. Diagnosing hypertension in Indigenous Canadians (DREAM-GLOBAL): A randomized controlled trial to compare the effectiveness of short message service messaging for management of hypertension: Main results. J Clin Hypertens (Greenwich) 2019 Jan;21(1):29-36 [FREE Full text] [doi: 10.1111/jch.13434] [Medline: 30474909]
38. Barsky J, Hunter R, Mcallister C, Yeates K, Campbell N, Liu P, et al. Evaluation of Implementation, User Perspectives and Feedback from DREAM-GLOBAL, a mobile Health Intervention in Canadian Indigenous and Tanzanian Communities for Individuals living with Hypertension. JMIR mHealth and uHealth 2019:1-19 (forthcoming). [doi: 10.2196/preprints.12639]
39. Maar MA, Yeates K, Toth Z, Barron M, Boesch L, Hua-Stewart D, et al. Unpacking the Black Box: A Formative Research Approach to the Development of Theory-Driven, Evidence-Based, and Culturally Safe Text Messages in Mobile Health Interventions. JMIR Mhealth Uhealth 2016 Jan 22;4(1):e10 [FREE Full text] [doi: 10.2196/mhealth.4994] [Medline: 26800712]
40. Ramsden IM. Cultural Safety and Nursing Education in Aotearoa and Te Waipounamu. Wellington: Victoria University of Wellington; 2002. URL: https://www.nzno.org.nz/Portals/0/Files/Documents/Services/Library/2002%20RAMSDEN%20I%20Cultural%20Safety_Full.pdf [accessed 2019-03-01] [WebCite Cache ID 76cmFecGt]
41. Cameron M, Andersson N, McDowell I, Ledogar RJ. Culturally Safe Epidemiology: Oxymoron or Scientific Imperative. Pimatisiwin 2010;8(2):89-116 [FREE Full text] [Medline: 20975852]
42. Greenwood M, Lindsay N, King J, Loewen D. Ethical spaces and places: Indigenous cultural safety in British Columbia health care. AlterNative: An International Journal of Indigenous Peoples 2017 Jun 20;13(3):179-189 [FREE Full text] [doi: 10.1177/1177180117714411]
43. Wilson D, Neville S. Culturally safe research with vulnerable populations. Contemp Nurse 2009 Aug;33(1):69-79. [Medline: 19715497]
44. Maar MA, Yeates K, Perkins N, Boesch L, Hua-Stewart D, Liu P, et al. A Framework for the Study of Complex mHealth Interventions in Diverse Cultural Settings. JMIR Mhealth Uhealth 2017 Apr 20;5(4):e47 [FREE Full text] [doi: 10.2196/mhealth.7044] [Medline: 28428165]
45. Government of Canada, Canadian Institutes of Health Research. CIHR Guidelines for Health Research Involving Indigenous People (2007-2010). 2013. URL: http://www.cihr-irsc.gc.ca/e/29134.html [accessed 2019-03-04]

46. Tobias JK, Richmond CAM, Luginaah I. Community-based participatory research (CBPR) with indigenous communities: producing respectful and reciprocal research. J Empir Res Hum Res Ethics 2013 Apr;8(2):129-140. [doi: 10.1525/jer.2013.8.2.129] [Medline: 23651937]

Abbreviations

BP: blood pressure
CHR: community health representative
DREAM-GLOBAL: Diagnosing Hypertension-Engaging Action and Management in Getting Lower Blood Pressure in Indigenous Peoples and Low- and Middle- Income Countries
eHealth: electronic health
I-RREACH: Intervention and Research Readiness Engagement and Assessment of Community Health Care
mHealth: mobile health
PI: principal investigator
RCT: randomized controlled trial
SMS: short message service

Edited by G Eysenbach; submitted 01.04.19; peer-reviewed by A Quinn, A Wright, C Vis; comments to author 06.09.19; revised version received 11.09.19; accepted 29.09.19; published 04.11.19.

Please cite as:
Maar MA, Beaudin V, Yeates K, Boesch L, Liu P, Madjedi K, Perkins N, Hua-Stewart D, Beaudin F, Wabano MJ, Tobe SW
Wise Practices for Cultural Safety in Electronic Health Research and Clinical Trials With Indigenous People: Secondary Analysis of a Randomized Clinical Trial
J Med Internet Res 2019;21(11):e14203
URL: https://www.jmir.org/2019/11/e14203
doi:10.2196/14203
PMID:31682574

Original Paper

Validation of an Independent Web-Based Tool for Measuring Visual Acuity and Refractive Error (the Manifest versus Online Refractive Evaluation Trial): Prospective Open-Label Noninferiority Clinical Trial.

Robert P L Wisse[1*], MD, PhD; Marc B Muijzer[1*], BSc; Francesco Cassano[2], MSc; Daniel A Godefrooij[1], MSc, MD, PhD; Yves F D M Prevoo[2], DVM, MBA; Nienke Soeters[1], PhD

[1]Utrecht Cornea Research Group, Ophthalmology Department, University Medical Center Utrecht, Utrecht, Netherlands

[2]Easee BV, Amsterdam, Netherlands

[*]these authors contributed equally

Corresponding Author:
Robert P L Wisse, MD, PhD
Utrecht Cornea Research Group
Ophthalmology Department
University Medical Center Utrecht
Utrecht
Netherlands
Phone: 31 88 75 51683
Email: r.p.l.wisse@umcutrecht.nl

Abstract

Background: Digital tools provide a unique opportunity to increase access to eye care. We developed a Web-based test that measures visual acuity and both spherical and cylindrical refractive errors. This test is Conformité Européenne marked and available on the Easee website. The purpose of this study was to compare the efficacy of this Web-based tool with traditional subjective manifest refraction in a prospective open-label noninferiority clinical trial.

Objective: The aim of this study was to evaluate the outcome of a Web-based refraction compared with a manifest refraction (golden standard).

Methods: Healthy volunteers from 18 to 40 years of age, with a refraction error between –6 and +4 diopter (D), were eligible. Each participant performed the Web-based test, and the reference test was performed by an optometrist. An absolute difference in refractive error of <0.5 D was considered noninferior. Reliability was assessed by using an intraclass correlation coefficient (ICC). Both uncorrected and corrected visual acuity were measured.

Results: A total of 200 eyes in 100 healthy volunteers were examined. The Web-based assessment of refractive error had excellent correlation with the reference test (ICC=0.92) and was considered noninferior to the reference test. Uncorrected visual acuity was similar with the Web-based test and the reference test (*P*=.21). Visual acuity was significantly improved using the prescription obtained by using the Web-based tool (*P*<.01). The Web-based test provided the best results in participants with mild myopia (ie, <3 D), with a mean difference of 0.02 (SD 0.49) D (*P*=.48) and yielding a corrected visual acuity of >1.0 in 90% (n=77) of participants.

Conclusions: Our results indicate that Web-based eye testing is a valid and safe method for measuring visual acuity and refractive error in healthy eyes, particularly for mild myopia. This tool can be used for screening purposes, and it is an easily accessible alternative to the subjective manifest refraction test.

Trial Registration: Clinicaltrials.gov NCT03313921; https://clinicaltrials.gov/ct2/show/NCT03313921.

(J Med Internet Res 2019;21(11):e14808) doi:10.2196/14808

KEYWORDS

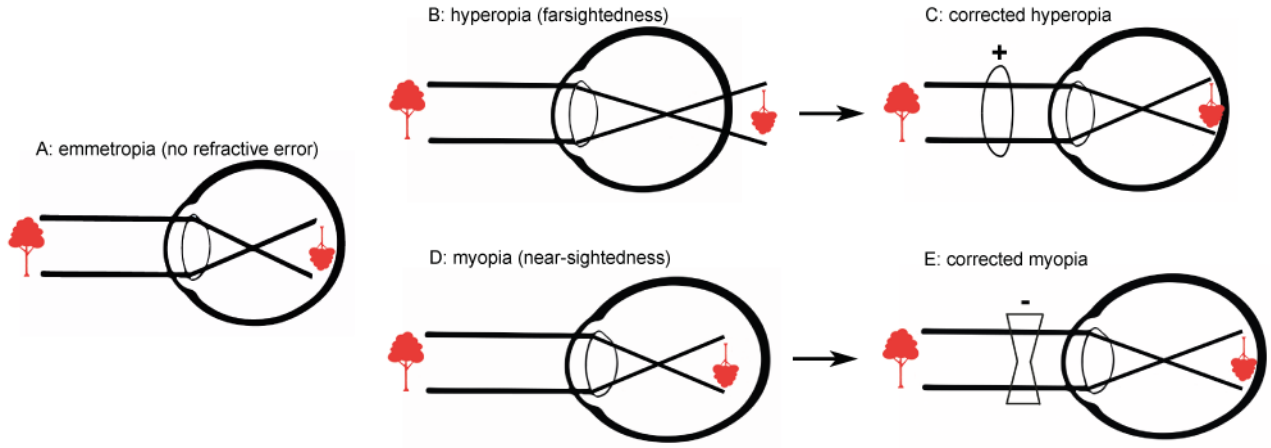
digital refraction; easee; telemedicine; medical informatics; refractive error

Introduction

Background

Globally, approximately 60% of individuals require a visual aid, such as spectacles or contact lenses, for proper visual acuity [1,2]. Moreover, studies have shown that the incidence of myopia (ie, nearsightedness) is increasing steadily because of higher literacy rates and increasing urbanization [3,4]. The World Health Organization has reported that these so-called refractive errors—if not corrected—represent the principal cause of visual impairment [5]. Strikingly, nearly 50% of preventable visual impairment is caused by the use of inappropriate spectacles or lenses, with severe economic implications [3,6]. Even in countries with readily accessible health care services, this rate remains unacceptably high, and calls for a new way of thinking about how visual aids are prescribed [2]. To improve eye health in our global population, we need access to reliable, affordable tools for measuring refractive error. In today's digital era, the ability to digitize the refractive exam is the logical solution. Indeed, many examples are available, supporting the robust potential of digital medicine, as well as the cultural and operational hurdles that must be overcome to bring medicine into the data-driven age [7,8]. To increase access to refractive testing, the Dutch company Easee BV in Amsterdam, the Netherlands, developed an algorithm-based Web-based tool that measures the refractive state of the eye by using a smartphone and computer screen. This tool is *Conformité Européenne* (CE) marked, complies with all required International Organization for Standardization (ISO) standards, and is currently available on the Web. Notwithstanding the apparent accessibility of this service, its validity and safety need to be studied and reported as a means to keep developers accountable for their health innovations. Traditionally, refractive error is measured by an eye care professional, in which trial lenses of various corrective strength are tested on the basis of the patient's responses, whereas a letter chart is used to assess the resulting visual acuity. The outcome of this test can include emmetropia (no refractive error), hyperopia (farsightedness), or myopia (nearsightedness), as well as astigmatism (a cylindrical error; see Figure 1). This so-called subjective manifest refraction test is currently considered the gold standard [9,10]. However, the quality of the measurement can depend upon a variety of factors, including the availability of the necessary equipment, a suitable environment for testing, the patient's ability and willingness to cooperate with the examiner, and the examiner's experience and training. Alternatively, refraction can be measured by using an automated approach, for example, with an automated refractor [11,12], an aberrometer [13], or adaptive optics [14]. Nevertheless, both the subjective and automated techniques require expensive medical equipment and qualified personnel, which can limit their availability and accessibility. Current developments in new refractive methods are summarized the Research in Context panel (Multimedia Appendix 1).

Figure 1. Optics of the eye. A: With no refractive error, the image is focused properly on the retina, providing perfect uncorrected visual acuity. B and D: In hyperopia (far-sightedness; B) and myopia (nearsightedness; D) the image falls either behind or in front of the retina, respectively. C and E: Lenses can be used to re-focus the image on the retina, restoring visual acuity.



Objectives

In this paper, we present the results of the Manifest versus Online Refraction Evaluation (MORE) trial, a study designed to validate this Web-based refractive assessment by comparing the outcome between the Web-based test and the subjective manifest refraction test, focusing on corrected visual acuity achieved by using the prescription obtained from the Web-based test.

Methods

Study Design and Recruitment

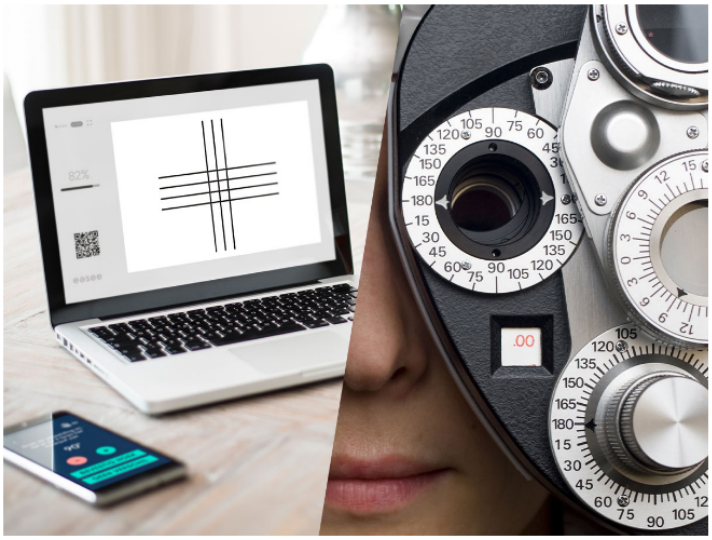
Data were prospectively collected in the open-label single-center noninferiority MORE trial, performed at the University Medical Center Utrecht in Utrecht, the Netherlands. The participants were healthy volunteers, from 18 to 40 years of age, with no history of eye disease or current evidence of eye disease. We excluded subjects whose refractive error was worse than –6 diopter (D; for myopia) or +4 D (for hyperopia) and subjects who had diabetes, were pregnant or lactated, or were unable to perform the Web-based test. All participants provided written

informed consent. All subjects underwent 3 consecutive tests designed to determine the refractive state of both eyes in the following order. First, the subject performed the index test using the Web-based refractive assessment tool with the Easee algorithm. Second, the refractive error was measured using autorefraction (Topcon RM 8800). Finally, an optometrist performed the reference test (manifest subjective refraction). The subject was blinded for the outcome of all tests. The subject's uncorrected distance visual acuity (UDVA) was recorded using a traditional Early Treatment Diabetic Retinopathy Study visual acuity chart and the Easee Web-based visual acuity test. Corrected distance visual acuity (CDVA) was measured using correction on the basis of the results of the manifest and Web-based refraction tests. Visual acuity was tested in accordance with ISO 8596, with regard to optotypes and room illumination [15]. The projected optotypes were randomized to mitigate any possible test-retest effect. Clinical agreement between manifest subjective refraction and autorefraction is generally considered excellent [10]; therefore, CDVA was not assessed using the results of the autorefraction test. The following data were recorded for each participant/eye: age, gender, laterality, medical history, previous prescription (if known), use of spectacles or contact lenses, UDVA, CDVA, and refractive outcome, including spherical and cylindrical power (in D) and axis (in degrees), which were converted into power vectors, using a Fourier analysis [16,17]. All procedures were performed in accordance with the Declaration of Helsinki, local and national laws regarding research (ie, the Act on Scientific Research Involving Humans), European directives with respect to privacy (General Data Protection Regulation 2016/679) and medical devices (Medical Device Regulation 2017/745), and the 2015 Standards for Reporting Diagnostic Accuracy Studies [18]. The study protocol was approved by our institution's Ethics Review Board (METC number: 17-524), and it was registered on the Web at clinicaltrials.gov (number: NCT03313921) and CCMO.nl (number: NL61478.041.17).

Information Regarding the Web-Based Tool

The Web-based tool for measuring refractive error uses a smartphone and a standard computer screen (Figure 2). This commercially available test is available via the website of Easee, and it uses the same algorithm described in this study; an 80-second video tutorial is also available at the website, and a clinical test flow is provided in the supplementary files. In brief, a smartphone functions as a remote control by which the user submits input from a distance of 3 m or 1.5 m to a computer screen that displays the Web-based test. Audio instructions (currently available in Dutch, English, and German) guide the user through the test, during which both eyes are tested consecutively. During the test, the user is presented a sequence of images and optotypes that the user must correctly identify, in addition to various grate sizes and astigmatism dials used to assess the cylindrical error. Any visual acuity below 1.0 (ie, worse than 20/20) is considered to be because of a refractive error. The direction of the refractive error (ie, hyperopia + or myopia –) is based on an adapted red/green duochrome test [19] and a questionnaire designed to discern between nearsightedness and farsightedness. A version of the Easee Web tool was custom built for this clinical trial in which only anonymized data were captured by the tool. The Web tool is classified as a class 1 medical device, which is in accordance with Medical Device Regulation 2017/745, and the software is classified as class A, which is in accordance with IEC 62304:2014.

Figure 2. An impression of the online refraction exam and its comparator the manifest refraction.



Statistical Analysis

The primary study outcome was refractive error, measured using the Web-based tool, and refractive error compared with a subjective manifest refraction and autorefraction. Specifically, we analyzed the sign of the refractive error (+/–), spherical power, cylindrical power, and axis, which were converted into power vectors by using a Fourier analysis [16,17]. An intraclass correlation coefficient (ICC) among the various methods was also calculated [20]. Autorefraction measurements were primarily used to provide a context for the level of correlation between a subjective manifest refraction and the Web-based tool. The secondary study outcomes included UDVA and CDVA, measured using the prescriptions obtained using the Web-based tool and a subjective manifest refraction. UDVA and CDVA were converted to logarithm of the minimum angle

of resolution (logMAR) values for statistical analysis. Groups were compared by using the 2-tailed paired Student's *t* test, or Pearson chi-square test. In addition, a multivariable analysis, using a generalized estimates equation, was used to correct for bilaterality (both eyes of the same patient included), age, and sex. Differences with a *P* value <.05 were considered statistically significant. A stratification in outcomes was defined in the study protocol for myopic and hyperopic results, as the subjective measurement of these distinct refractive states is prone to particular errors. A difference in spherical equivalent (SEQ) >0.5 D between the 3 refraction methods was considered to reflect a clinically significant difference; thus, this constituted the threshold for noninferiority [21,22]. The power calculation was based on an intraclass correlation for the 3 different refraction methods, using the following formula in R: Sample size(p=0.70,p0=0,k=3,alpha=(0.05/12),tails=2,power=0.80, by="p", step=0.025).

Initially, 50 healthy subjects with 100 healthy eyes were scheduled to enter the study. An interim analysis indicated that the algorithm yielded more outlier measurements than anticipated, thereby skewing the results. A second-generation algorithm was therefore developed, using the clinical data acquired to date. An extension was requested for the trial, and it was granted by our institution's Ethics Review Board. Any incomplete data were imputed, except when it concerned missing data from the primary outcome. Data were analyzed using IBM SPSS v25.0 (IBM).

Results

Description of the Study Population

A total of 200 eyes from 100 healthy subjects were included in the study; 1 eye was excluded from analysis because of amblyopia (lazy eye). All subjects were enrolled in the study between December 28, 2017 and January 28, 2019. The clinical characteristics of the participants are summarized in Table 1. Most of the subjects (62%, n=62) were regular users of spectacles or contact lenses. A total of 4 subjects reported receiving previous treatment for an ophthalmic condition; in all 4 cases, the ophthalmic condition resolved without sequelae. A total of 11 subjects reported ocular complaints at the time of the measurements; 8 subjects reported blurred vision, and 3 subjects reported other complaints, such as floaters and dry eyes. The mean test duration was 22 (SD 10) min (range 5-58). No adverse events or complications were recorded during the trial.

Table 1. Clinical characteristics of the study population.

Clinical characteristics[a]	Total (N=100)	Web-based test algorithm		*P* value[b]
		1st generation (N=36)	2nd generation (N=64)	
Age (years), mean (SD)	25.4 (4.7)	25.3 (4.2)	25.5 (4.9)	.86
Sex (male), n (%)	47 (47)	23 (64)	24 (38)	.001
Current use of visual aids, n (%)	62 (62)	21 (58)	42 (6)	.58
Spectacles	60 (60)	21 (58)	27 (61)	.88
Contact lenses	22 (22)	7 (20)	15 (23)	.54
Previous ophthalmic treatment, n (%)	4 (4)	3 (8)	1 (2)	.12
Ocular complaints, n (%)	13 (13)	3 (8)	10 (16)	.34
Medication use, n (%)	13 (13)	2 (6)	11 (17)	.08
Refractive error[c] , n (%)				
Emmetropia	16 (8)	4 (6)	12 (9)	N/A[d]
Mild myopia	119 (60)	40 (56)	79 (62)	N/A
Severe myopia	32 (16)	13 (18)	19 (15)	N/A
Hyperopia	32 (16)	15 (21)	17 (13)	N/A
Total	199	72	127	.24

[a]Except where indicated otherwise, data are presented as n (%).

[b]Calculated using an independent samples Student *t* test or Pearson chi-square test.

[c]Mild myopia was defined as refractive error of –3 D or less; severe myopia was defined as refractive error worse than –3 D. Refractive error was determined on the basis of the spherical equivalent of the manifest refraction value, and it is reported for both eyes separately.

[d]Not applicable.

Intraclass Correlation Coefficients

As a first measure of the concordance among the 3 methods for assessing refractive error, we measured the ICC. For this analysis, we included only each participant's right eye and based our calculations on the SEQ. The overall ICC of all 3 measurements was 0.93 (95% CI 0.90-0.96), and the overall ICC for manifest refraction and Web-based refraction was 0.89 (95% CI 0.84-0.93). When only measurements taken with the second-generation algorithm are considered, the ICC improved to 0.92 (95% CI 0.86-0.95). The latter can be considered an

excellent agreement [20]. Analyses based on vectors rather than SEQs did not materially alter these findings.

Reliability of Web-Based Visual Acuity Testing

UDVA was measured by using both the Web-based test and a visual acuity wall chart. UDVA data for the Web-based test were imputed for 6 participants because of a technical recording error. Our analysis revealed that the Web-based test provided UDVA values that were similar to results obtained by using a chart, with mean values of 0.67 (SD 0.33) versus 0.69 (SD 0.37), respectively (LogMAR: 0.33 (SD 0.30) vs 0.39 (SD 0.39); *P*=.21). In addition, the overall ICC of this measurement (for each participant's right eye only) was 0.89 (95% CI 0.83-0.92).

Overall Outcome for Measuring Refractive Error With the Web-Based Test Versus the Reference Test

In the entire study group, refractive error between the Web-based refraction test and the reference test differed by –0.18 (SD 0.77) D for participants with myopia and 0.63 (SD 0.89) D for participants with hyperopia. With respect to the participants with myopia, this difference was within our *a priori* threshold for defining noninferiority (see the Multimedia Appendices 2-4). When we analyzed only the participants who were tested using the second-generation algorithm, the difference in SEQ was –0.13 (SD 0.62) D for patients with myopia and 0.50 (SD 0.81) D for patients with hyperopia, both of which are within our threshold for noninferiority margin (see Table 2). Similar results were obtained when we corrected for the confounding factors bilaterality, age, and sex (data not shown).

Figure 3 shows the difference of the Web-based test compared with the reference test and with respect to the noninferiority limit. As can be observed, a majority of measurements fall within the noninferiority limit, and almost all measurements fall within the 95% CI. In addition, we summarized the distribution of the differences in refractive outcome by using the Web-based test and manifest refraction. Figure 4 shows the individual refractive error data measured for each patient; note that the 6 patients for whom data were missing are not included in these graphs.

Table 2. Refractive error and visual acuity measured in the myopic and hyperopic participants (second-generation algorithm; N=121 eyes).

Refractive error and visual acuity	Manifest refraction[a]	Online refraction[a]	Difference	95% CI	*P* value[b]	GEE model[c]	
						Beta value	*P* value[c]
Emmetropic and myopic eyes (n=104)							
Power vector (Diopter)[d]	1.59 (1.50)	1.47 (1.27)	0.12	0.00-0.24	.04	1.11	<.001
J0 vector (Diopter)	0.09 (0.29)	–0.01 (0.22)	0.10	0.04-0.15	N/A[e]	N/A	N/A
J45 vector (Diopter)	0.01 (0.17)	–0.01 (0.15)	0.00	–0.04 to 0.03	N/A	N/A	N/A
Spherical equivalent (Diopter)	–1.54 (–1.52)	–1.41 (1.31)	0.13	–0.25 to –0.01	N/A	N/A	N/A
Spherical power (Diopter)	–1.31 (1.43)	–1.30 (1.31)	–0.01	–0.13 to 0.10	N/A	N/A	N/A
Cylindrical power (Diopter)	–0.45 (0.51)	–0.23 (0.47)	–0.22	–0.34 to –0.11	N/A	N/A	N/A
Cylindrical axis (degrees)	97 (58)	101 (51)	–4	–24 to 16	N/A	N/A	N/A
CDVA[f] logarithm of the minimum angle of resolution[g]	–0.14 (0.06)	–0.03 (0.18)	–0.11	–0.14 to –0.08	<.001	.08	.13
CDVA Snellen[h,i]	1.38 (0.20)	1.15 (0.35)	0.25	0.18-0.32	N/A	N/A	N/A
Hyperopic eyes (n=17)							
Power vector (Diopter)[d]	0.58 (0.45)	0.33 (0.48)	0.25	0.14-0.37	.001	.84	<.001
J0 vector (Diopter)	0.03 (0.21)	0.00 (0.14)	0.02	–0.11 to 0.15	N/A	N/A	N/A
J45 vector (Diopter)	–0.02 (0.17)	0.03 (0.09)	0.05	–0.15 to 0.05	N/A	N/A	N/A
Spherical equivalent (Diopter)	0.53 (0.44)	0.03 (0.57)	0.50	0.11-0.89	N/A	N/A	N/A
Spherical power (Diopter)	0.71 (0.57)	0.10 (0.58)	0.61	0.16-1.04	N/A	N/A	N/A
Cylindrical power (Diopter)	–0.35 (0.40)	–0.15 (0.29)	–0.21	–0.38 to –0.04	N/A	N/A	N/A
Cylindrical axis (degrees)	53 (50)	46 (40)	8	–72 to 88	N/A	N/A	N/A
CDVA logarithm of the minimum angle of resolution[h]	–0.13 (0.06)	–0.10 (0.11)	–0.03	–0.08 to 0.02	.20	.25	.54
CDVA Snellen[h,i]	1.37 (0.19)	1.29 (0.28)	0.08	–0.06 to 0.21	N/A	N/A	N/A

[a]Unless otherwise specified, reported as mean (SD).

[b]Paired-sample Student *t* test was performed for predefined primary and secondary outcome parameters only.

[c]Generalized estimates equation model to statistically correct for the inclusion of 2 eyes of one subject, age, and sex.

[d]Spherical and cylindrical power and axes were translated into vectors using Fourier analysis.

[e]Not applicable.

[f]CDVA: corrected distance visual acuity.

[g]Assessed with either the manifest or Web-based achieved correction.

[h]Snellen, decimal visual acuity.

[i]Statistical tests were performed only on predefined parameters (power vector for refraction and logarithm of the minimum angle of resolution for visual acuity).

Figure 3. The difference between the refractive error measurement of the first (red) and second generation (blue) online refraction test compared to the outcome of the manifest refraction with respect to the non-inferiority limit (green area) and 95% confidence interval (dashed lines).

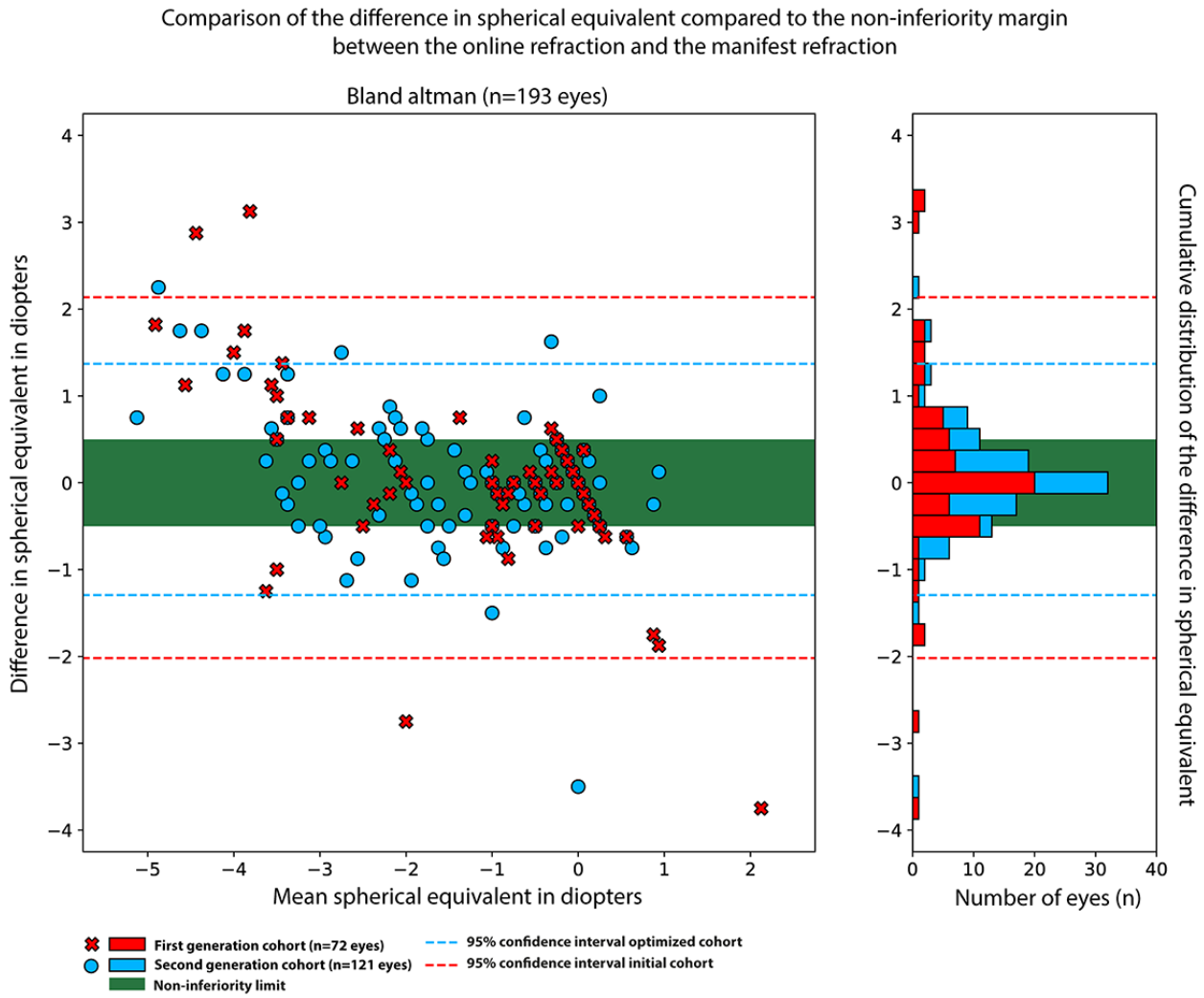
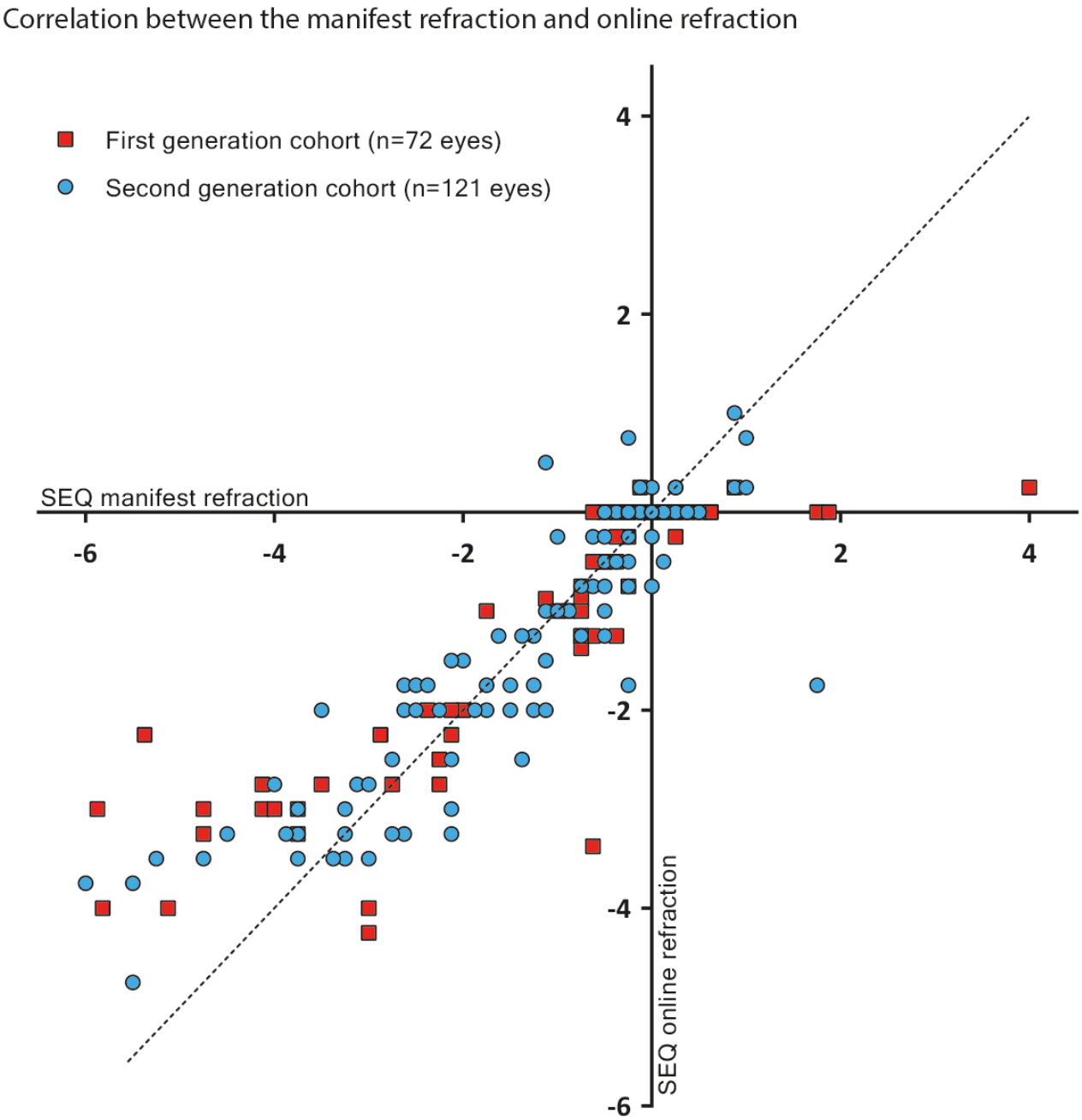


Figure 4. Refractive error measured using the online test was plotted against refractive error measured using manifest refraction; each symbol represents an individual eye measured in a participant who was tested using the first-generation algorithm (red squares) or the second-generation algorithm (blue circles). The 45° dashed line represents an ideal fit. Outliers are identified particularly in the high-myopia group (bottom-left), and these differences are reduced in the second generation cohort. SEQ: spherical equivalent.



Overall Visual Acuity Measured Using the Web-Based Refraction Test and Manifest Refraction

Visual acuity improved significantly using the prescription obtained by using the Web-based refraction test, particularly when using the second-generation algorithm. Specifically, the UDVA was 0.66 (SD 0.41) (LogMAR 0.32 [SD 0.40]), and it improved to a CDVA of 1.17±0.34 (LogMAR –0.04 [SD 0.17]; *P*<.01). Interestingly, we found that CDVA in the hyperopic participants did not differ significantly between the Web-based refraction test (1.29 [SD 0.28], LogMAR –0.10 [SD 0.11]) and the manifest refraction test (1.37 [SD 0.19], LogMAR –0.13 [SD 0.06]; *P*=.20). This is likely because of the accommodation reflex that corrects residual hyperopic refractive errors [1]. A multivariable Generalized Estimating Equations (GEE) analysis did not reveal any major confounders (Table 2).

For myopic participants, the visual acuity (CDVA) differed significantly between the Web-based refraction test (1.15 [SD 0.35]; LogMAR –0.03 [SD 0.18]) and the manifest refraction test (1.38 [SD 0.19]; LogMAR –0.14 [SD 0.06]; *P*<.01). Contrary to hyperopia, even a small an uncorrected residual myopic refractive error will negatively influence distance visual acuity [1]. A multivariable GEE analysis revealed that confounding factors influenced this difference, although with a very small effect size, attributable to the second-generation

cohort, harboring relatively more myopic females. Analysis of Web-based test meta-data revealed no clues to a difference in performance of male versus female participants.

The Ability of the Web-Based Refraction Test to Correctly Distinguish Myopia Versus Hyperopia

In nearly every case, the Web-based refraction was able to correctly determine the participant's refractive error as either myopia or hyperopia, with the exception of 4 cases. A total of 1 case fell within the noninferiority margin, with a difference of 0.25 D, and CDVA was similar for this eye when corrected with either prescription. The other 3 cases differed to a clinically relevant level: -1.125 versus + 0.50, +0.125 versus –0.50, and +1.75 versus –1.75 for the Web-based and manifest refraction test, respectively. In 195 of 199 cases (98%) of the Web-based assessments, the signation was correct.

Subgroup Analysis of Participants With Mild Myopia

A majority of eyes in our study were classified as having mild myopia, which is consistent with mild myopia being the most common refractive error in the general population [23]. A subgroup analysis was performed in the eyes, with a refractive error between –3 and 0 D, 91 eyes in total. Using the prescription obtained with the Web-based refraction test, the eyes with mild myopia had a markedly better CDVA compared with the entire group of eyes with myopia (1.22 [SD 0.29]; LogMAR –0.08 [SD 0.11]), with 90% (n=77) of the participants scoring over 1.0. The average difference in refractive error between the 2 tests is now reduced to 0.02 (SD 0.49) D (P=.48), and 80% of the Web-based refraction tests were within SD 0.5 D of the reference test. Notwithstanding, the manifest refraction test yielded a slightly better CDVA (1.39 [SD 0.20]; LogMAR –0.13 [SD 0.06]; P<.01). Outcomes are reported in detail in Table 3. In this subgroup, a GEE multivariable analysis indicated a comparable confounding effect as described earlier regarding the overall outcomes.

Table 3. Refractive error and visual acuity measured in the mildly myopic participants (second-generation algorithm; N=86 eyes).

Refractive error and visual acuity	Manifest refraction[a]	Web-based refraction[a]	Difference	95% CI	*P* value[b]	GEE model[c]	
						Beta value	*P* value[c]
Power vector (Diopter)[d]	1.04 (0.86)	1.07 (0.96)	0.03	–0.13 to 0.06	.48	.82	<.01
J0 vector (Diopter)	0.05 (0.24)	–0.00 (0.23)	0.05	0.00 to 0.10	N/A[f]	N/A	N/A
J45 vector (Diopter)	–0.14 (0.14)	–0.10 (0.16)	–0.04	–0.04 to 0.03	N/A	N/A	N/A
Spherical equivalent (Diopter)	–0.98 (0.88)	–1.00 (1.00)	0.02	–0.09 to 0.13	N/A	N/A	N/A
Spherical power (Diopter)	–0.78 (0.85)	–0.87 (0.96)	0.09	–0.03 to 0.20	N/A	N/A	N/A
Cylindrical power (Diopter)	–0.38 (0.41)	–0.26 (0.50)	0.12	–0.24 to –0.02	N/A	N/A	N/A
Cylindrical axis (degrees)	101 (59)	105 (50)	4	–27 to 19	N/A	N/A	N/A
Corrected distance visual acuity logarithm of the minimum angle of resolution[e,g]	–0.14 (0.06)	–0.08 (0.10)	–0.06	–0.09 to –0.04	<.01	.15	.03
Corrected distance visual acuity Snellen[e,g,h]	1.39 (0.20)	1.22 (0.29)	0.17	0.11 to 0.23	N/A	N/A	N/A

[a]Unless otherwise specified, reported as mean (SD).

[b]Paired-sample *t* test performed for predefined primary and secondary outcome parameter only.

[c]Generalized Estimates Equation model to statistically correct for the inclusion of 2 eyes of one subject, age, and sex.

[d]Spherical and cylindrical power and axes were translated in vectors by Fourier analysis.

[e]Assessed with either the manifest or Web-based achieved correction.

[f]Not applicable.

[g]Snellen: decimal visual acuity.

[h]Statistical tests were performed only on predefined parameters (power vector for refraction and logarithm of the minimum angle of resolution for visual acuity).

Discussion

Principal Findings

In this noninferiority clinical trial, we compared a Web-based tool for measuring refractive error with the current gold standard, the subjective manifest refraction. Our analysis revealed excellent correlations between the 2 tests (ICC 0.92). Thus, we conclude that the Web-based test can be considered noninferior to manifest refraction. Importantly, the Web-based test provided a reliable measure visual acuity, similar to using a traditional wall chart, and visual acuity improved significantly using the prescription obtained by using the Web-based test, particularly among participants with mild myopia. This study provides the necessary validity and safety data for the Web-based eye test offered by Easee.

The Web-based tool measures the eye's visual acuity and translates this outcome into refractive error, while assuming that any error is caused solely by an uncorrected refractive error. Thus, patients with a vision-limiting eye condition, such as amblyopia, cataract, or a retinal disease, may not necessarily obtain a reliable measure of refractive error by using the Web-based tool. In practice, this effect is mitigated by including a disclaimer for patients who have such an eye condition, although the patients must be aware of having such a condition to heed this disclaimer. It is also important to note that refractive errors in subjects with a high visual acuity or eye conditions that do not limit vision (eg, glaucoma or mild diabetic retinopathy) will likely not be detected by this Web-based test.

Considerations

However, some limitations of the study itself should be taken into consideration. No randomization of the test order was performed and could have impacted our results. Subjects may become tired during the assessments. Although because of the fixed test order, this should have impacted all subject similarly. We consider the learning or training effect during the tests as negligible. The 3 methods of refractive assessment are very different, and randomized projected optotypes were used to assess visual acuity. In addition, subjects were blinded for the outcome to prevent testing bias. Notwithstanding, the observer had access to the test outcomes; thus, an observer bias cannot fully be excluded. The Web-based test has been validated for use in healthy individuals. Further studies should be performed to test the feasibility of using this Web-based tool in children and populations with a higher incidence of eye disease. Another consideration is the role of accommodation during the test, which is defined as a semivoluntary reflex, causing the eye to focus on a nearby object; this reflex can increase the eye's refractive power and can therefore mask a residual hyperopic refractive error. We found that the Web-based test tended to underestimate a hyperopic refractive error by an average of 0.5 D, which suggests that the accommodation reflex may have played a role in these participants. We consider the manifest refraction test a more powerful tool to measure the full hyperopic refractive error. Nevertheless, undercorrecting a hyperopic refractive error may be preferred over issuing the full-strength prescription, and this can sufficiently alleviate the patient's visual complaints [24]. All measurements were performed in accordance with ISO standards regarding visual acuity testing, revealing that a fully autonomous algorithm is capable of nearly matching the results obtained by an optometrist, at least in a healthy population. In daily practice, not all refraction assessments are performed by an optometrist, and not all assessments are performed under ideal conditions. Depending on local regulations and customs, a technician or a trained optician may perform the exam. Moreover, an authoritative consumer report revealed that prescriptions issued by eye care professionals can have wide variability [25]. Importantly, although our Web-based refraction test depends on the patient's input, it has zero variability with respect to interpreting the patient's responses, and it should provide high test-retest reproducibility. Further research is needed to determine whether the Web-based tool has high intrasubject consistency.

Practical Perspective

The recent increase in digitization has increased the availability and accessibility of the Web-based refraction test, as anyone with a laptop and smartphone can complete the test without the need to visit an eye care professional. Moreover, 2.7 billion people are estimated to have a smartphone in 2019 [26], and approximately 97% of our target patient population—users from 18 to 45 years of age—have a smartphone [27]. The availability of a Web-based refraction fits into the current trend of digitalization, and this provides consumers with more flexibility in planning their eye test. In addition, the Web-based refraction benefits patients in areas with limited access to eye care professionals. Basatwrous et al have convincingly shown that creating a comprehensive digital eye care ecosystem can elevate the overall health in a rural community [28]. The Research in Context panel summarizes current initiatives on remote eye testing. A future perspective is the use of the Easee eye test in a clinical environment with automated data entry in the electronic health record, as well as integration in clinical care, for example, cataract, macular degeneration, and glaucoma patients. The measurements provided by the Web-based refraction test were not subjected to post hoc processing, and these were entered directly into the database for analysis. Despite the high rate of concordance between the Web-based refraction test and the manifest refraction test, the Web-based test was not able to detect all outliers and unusual results. Therefore, additional interpretation of previous prescriptions and remote validation of the data by a qualified optometrist may still be warranted. Importantly, the Web-based refraction test is not designed to fully replace a comprehensive eye exam by a trained eye care professional, and users must comply with the test's terms and conditions; failure to do so will prompt the advice to visit an eye care professional.

Conclusions

Here, we report that the Easee Web-based test for measuring refractive error provides a safe, valid method for obtaining a corrective prescription in individuals with healthy eyes, particularly patients with mild myopia. Using the prescription obtained with the Web-based test significantly improves visual acuity to a degree similar to the prescription obtained using manifest refraction. Therefore, the Web-based refraction test

provides a user-friendly, easily accessible alternative to the traditional subjective manifest refraction test, although it should not be considered a replacement for a comprehensive eye examination. The Web-based test is CE marked; therefore, it meets the requirements established by the European Union with respect to safety and health.

Acknowledgments

The study was initiated by the University Medical Center Utrecht, The Netherlands, and funded by Easee BV, Amsterdam, The Netherlands. The authors are grateful to Dr Paul Westers of the Julius Center for Health Sciences, Department of Biostatistics and Research Support, University Medical Center Utrecht for providing an independent analysis of the data reported in this manuscript, and to Emma van Hamersveld BSc, for her assistance in composing the regulatory and ethical board review documentation.

Conflicts of Interest

YP is founder and shareholder of Easee BV. FC is employed by Easee BV. RW is employed by the UMC Utrecht and medical advisor and shareholder of Easee BV. MBM is employed by the UMC Utrecht and a consultant for Easee BV. NS and DG have no potential conflicts of interest, and they supervised the medical and statistical contents of the manuscript.

Multimedia Appendix 1
Panel - Research in context.
[PDF File (Adobe PDF File), 237 KB - jmir_v21i11e14808_app1.pdf]

Multimedia Appendix 2
Clinical test flow second generation online refraction test.
[PDF File (Adobe PDF File), 1932 KB - jmir_v21i11e14808_app2.pdf]

Multimedia Appendix 3
Clinical test flow second generation online refraction test.
[PDF File (Adobe PDF File), 7573 KB - jmir_v21i11e14808_app3.pdf]

Multimedia Appendix 4
Supplementary data overall results.
[PDF File (Adobe PDF File), 250 KB - jmir_v21i11e14808_app4.pdf]

References

1. Vitale S, Ellwein L, Cotch MF, Ferris 3rd FL, Sperduto R. Prevalence of refractive error in the United States, 1999-2004. Arch Ophthalmol 2008 Aug;126(8):1111-1119 [FREE Full text] [doi: 10.1001/archopht.126.8.1111] [Medline: 18695106]
2. Pascolini D, Mariotti SP. Global estimates of visual impairment: 2010. Br J Ophthalmol 2012 May;96(5):614-618. [doi: 10.1136/bjophthalmol-2011-300539] [Medline: 22133988]
3. Williams KM, Bertelsen G, Cumberland P, Wolfram C, Verhoeven VJ, Anastasopoulos E, European Eye Epidemiology (E(3)) Consortium. Increasing prevalence of myopia in Europe and the impact of education. Ophthalmology 2015 Jul;122(7):1489-1497 [FREE Full text] [doi: 10.1016/j.ophtha.2015.03.018] [Medline: 25983215]
4. Pan CW, Ramamurthy D, Saw SM. Worldwide prevalence and risk factors for myopia. Ophthalmic Physiol Opt 2012 Jan;32(1):3-16. [doi: 10.1111/j.1475-1313.2011.00884.x] [Medline: 22150586]
5. Resnikoff S, Pascolini D, Mariotti SP, Pokharel GP. Global magnitude of visual impairment caused by uncorrected refractive errors in 2004. Bull World Health Organ 2008 Jan;86(1):63-70 [FREE Full text] [doi: 10.2471/blt.07.041210] [Medline: 18235892]
6. Vitale S, Cotch MF, Sperduto R, Ellwein L. Costs of refractive correction of distance vision impairment in the United States, 1999-2002. Ophthalmology 2006 Dec;113(12):2163-2170. [doi: 10.1016/j.ophtha.2006.06.033] [Medline: 16996610]
7. de Fauw J, Ledsam JR, Romera-Paredes B, Nikolov S, Tomasev N, Blackwell S, et al. Clinically applicable deep learning for diagnosis and referral in retinal disease. Nat Med 2018 Sep;24(9):1342-1350. [doi: 10.1038/s41591-018-0107-6] [Medline: 30104768]
8. Norgeot B, Glicksberg BS, Butte AJ. A call for deep-learning healthcare. Nat Med 2019 Jan;25(1):14-15. [doi: 10.1038/s41591-018-0320-3] [Medline: 30617337]
9. Elliott DB. What is the appropriate gold standard test for refractive error? Ophthalmic Physiol Opt 2017 Mar;37(2):115-117. [doi: 10.1111/opo.12360] [Medline: 28211178]

10. Bullimore MA, Fusaro RE, Adams CW. The repeatability of automated and clinician refraction. Optom Vis Sci 1998 Aug;75(8):617-622. [doi: 10.1097/00006324-199808000-00028] [Medline: 9734807]
11. Wood IC. A review of autorefractors. Eye (Lond) 1987;1(Pt 4):529-535. [doi: 10.1038/eye.1987.80] [Medline: 3443207]
12. Pesudovs K, Weisinger HS. A comparison of autorefractor performance. Optom Vis Sci 2004 Jul;81(7):554-558. [doi: 10.1097/00006324-200407000-00018] [Medline: 15252356]
13. Nguyen MT, Berntsen DA. Aberrometry repeatability and agreement with autorefraction. Optom Vis Sci 2017 Sep;94(9):886-893 [FREE Full text] [doi: 10.1097/OPX.0000000000001107] [Medline: 28727613]
14. Hervella L, Villegas EA, Prieto PM, Artal P. Assessment of subjective refraction with a clinical adaptive optics visual simulator. J Cataract Refract Surg 2019 Jan;45(1):87-93 [FREE Full text] [doi: 10.1016/j.jcrs.2018.08.022] [Medline: 30309774]
15. ISO - International Organization for Standardization. 2017. Ophthalmic Optics -- Visual Acuity Testing -- Standard and Clinical Optotypes and Their Presentation URL: https://www.iso.org/standard/69042.html [accessed 2019-09-12]
16. Raasch T. Clinical refraction in three-dimensional dioptric space revisited. Optom Vis Sci 1997 Jun;74(6):376-380. [doi: 10.1097/00006324-199706000-00020] [Medline: 9255815]
17. Thibos LN, Wheeler W, Horner D. Power vectors: an application of Fourier analysis to the description and statistical analysis of refractive error. Optom Vis Sci 1997 Jun;74(6):367-375. [doi: 10.1097/00006324-199706000-00019] [Medline: 9255814]
18. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, STARD Group. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. Br Med J 2015 Oct 28;351:h5527 [FREE Full text] [doi: 10.1136/bmj.h5527] [Medline: 26511519]
19. Yanof M, Duker JS. Ophthalmology: Expert Consult. Third Edition. Maryland Heights: Mosby Elsevier; 2009.
20. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016 Jun;15(2):155-163. [doi: 10.1016/j.jcm.2016.02.012] [Medline: 27330520]
21. Goss DA, Grosvenor T. Reliability of refraction--a literature review. J Am Optom Assoc 1996 Oct;67(10):619-630. [Medline: 8942135]
22. Macaya F, Ryan N, Salinas P, Pocock SJ. Challenges in the design and interpretation of noninferiority trials: insights from recent stent trials. J Am Coll Cardiol 2017 Aug 15;70(7):894-903 [FREE Full text] [doi: 10.1016/j.jacc.2017.06.039] [Medline: 28797360]
23. Tideman JW, Snabel MC, Tedja MS, van Rijn GA, Wong KT, Kuijpers RW, et al. Association of axial length with risk of uncorrectable visual impairment for Europeans with myopia. JAMA Ophthalmol 2016 Dec 1;134(12):1355-1363. [doi: 10.1001/jamaophthalmol.2016.4009] [Medline: 27768171]
24. Park KA, Kim SM, Oh SY. The maximal tolerable reduction in hyperopic correction in patients with refractive accommodative esotropia: a 6-month follow-up study. Am J Ophthalmol 2011 Mar;151(3):535-41.e2. [doi: 10.1016/j.ajo.2010.09.023] [Medline: 21236414]
25. van der Staak B. Consumentenbond. 2018. Opticiens Vergeleken: Resultaten Oogmetingen Heel Wisselend URL: https://www.consumentenbond.nl/nieuws/2018/opticiens-vergeleken-resultaten-oogmetingen-heel-wisselend [accessed 2019-03-01]
26. Cisco. 2019. Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2017–2022 URL: https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white-paper-c11-738429.html [accessed 2019-09-12]
27. Deloitte US | Audit, Consulting, Advisory, and Tax Services. 2013. Global Mobile Consumer Survey 2013 - Addicted to Connectivity: The Dutch Mobile Consumer Perspectiv URL: https://www2.deloitte.com/content/dam/Deloitte/nl/Documents/technology-media-telecommunications/deloitte-nl-global-mobile-consumer-survey%20(1).pdf [accessed 2019-09-12]
28. Bastawrous A, Rono HK, Livingstone IA, Weiss HA, Jordan S, Kuper H, et al. Development and validation of a smartphone-based visual acuity test (Peek Acuity) for clinical practice and community-based fieldwork. JAMA Ophthalmol 2015 Aug;133(8):930-937 [FREE Full text] [doi: 10.1001/jamaophthalmol.2015.1468] [Medline: 26022921]

Abbreviations

CDVA: corrected distance visual acuity
CE: Conformité Européenne
D: diopter
GEE: generalized estimating equations
ICC: intraclass correlation coefficient
ISO: International Organization for Standardization
LogMAR: logarithm of the minimum angle of resolution
MORE: Manifest versus Online Refraction Evaluation
SEQ: spherical equivalent
UDVA: uncorrected distance visual acuity

Edited by G Eysenbach; submitted 29.05.19; peer-reviewed by A Lutz de Araujo, H Yang; comments to author 03.07.19; revised version received 15.07.19; accepted 17.08.19; published 08.11.19.

Please cite as:
Wisse RPL, Muijzer MB, Cassano F, Godefrooij DA, Prevoo YFDM, Soeters N
Validation of an Independent Web-Based Tool for Measuring Visual Acuity and Refractive Error (the Manifest versus Online Refractive Evaluation Trial): Prospective Open-Label Noninferiority Clinical Trial.
J Med Internet Res 2019;21(11):e14808
URL: https://www.jmir.org/2019/11/e14808
doi:10.2196/14808
PMID:31702560

Original Paper

The Service of Research Analytics to Optimize Digital Health Evidence Generation: Multilevel Case Study

Quynh Pham[1,2], PhD; James Shaw[1,3], PhD; Plinio P Morita[1,4], PEng, PhD; Emily Seto[1,2], PEng, PhD; Jennifer N Stinson[1,5,6,7], CPNP, RN, PhD; Joseph A Cafazzo[1,2,8], PEng, PhD

[1]Institute of Health Policy, Management, and Evaluation, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada

[2]Centre for Global eHealth Innovation, Techna Institute, University Health Network, Toronto, ON, Canada

[3]Women's College Hospital, Institute for Health System Solutions and Virtual Care, Toronto, ON, Canada

[4]School of Public Health and Health Systems, Faculty of Applied Health Sciences, University of Waterloo, Waterloo, ON, Canada

[5]Department of Anesthesia and Pain Medicine, The Hospital for Sick Children, Toronto, ON, Canada

[6]Lawrence S Bloomberg Faculty of Nursing, University of Toronto, Toronto, ON, Canada

[7]Child Health Evaluative Sciences Research Institute, The Hospital for Sick Children, Toronto, ON, Canada

[8]Institute of Biomaterials and Biomedical Engineering, Faculty of Applied Science and Engineering, University of Toronto, Toronto, ON, Canada

Corresponding Author:
Quynh Pham, PhD
Institute of Health Policy, Management, and Evaluation
Dalla Lana School of Public Health
University of Toronto
Health Sciences Building, Suite 425
155 College Street
Toronto, ON, M5T 3M6
Canada
Phone: 1 416 340 4800 ext 4765
Email: q.pham@mail.utoronto.ca

Abstract

Background: The widespread adoption of digital health interventions for chronic disease self-management has catalyzed a paradigm shift in the selection of methodologies used to evidence them. Recently, the application of digital health research analytics has emerged as an efficient approach to evaluate these data-rich interventions. However, there is a growing mismatch between the promising evidence base emerging from analytics mediated trials and the complexity of introducing these novel research methods into evaluative practice.

Objective: This study aimed to generate transferable insights into the process of implementing research analytics to evaluate digital health interventions. We sought to answer the following two research questions: (1) how should the service of research analytics be designed to optimize digital health evidence generation? and (2) what are the challenges and opportunities to scale, spread, and sustain this service in evaluative practice?

Methods: We conducted a qualitative multilevel embedded single case study of implementing research analytics in evaluative practice that comprised a review of the policy and regulatory climate in Ontario (macro level), a field study of introducing a digital health analytics platform into evaluative practice (meso level), and interviews with digital health innovators on their perceptions of analytics and evaluation (microlevel).

Results: The practice of research analytics is an efficient and effective means of supporting digital health evidence generation. The introduction of a research analytics platform to evaluate effective engagement with digital health interventions into a busy research lab was ultimately accepted by research staff, became routinized in their evaluative practice, and optimized their existing mechanisms of log data analysis and interpretation. The capacity for research analytics to optimize digital health evaluations is highest when there is (1) a collaborative working relationship between research client and analytics service provider, (2) a data-driven research agenda, (3) a robust data infrastructure with clear documentation of analytic tags, (4) in-house software development expertise, and (5) a collective tolerance for methodological change.

Conclusions: Scientific methods and practices that can facilitate the agile trials needed to iterate and improve digital health interventions warrant continued implementation. The service of research analytics may help to accelerate the pace of digital health evidence generation and build a data-rich research infrastructure that enables continuous learning and evaluation.

(J Med Internet Res 2019;21(11):e14849) doi:10.2196/14849

KEYWORDS

research analytics; effective engagement; digital health; mobile health; implementation; log data; service design; chronic disease

Introduction

Background

The widespread adoption of digital health interventions for chronic disease self-management has catalyzed a paradigm shift in the selection of methodologies used to evidence them [1]. This turn toward alternative research designs and methods is predicated on the understanding that traditional approaches to evidence generation cannot keep pace with digital health innovation [2]. Transformative advances to the technological components powering these novel interventions have rapidly changed their capacity to improve health outcomes [3]. As such, they demand agile and iterative evaluations that can continuously measure their effects in real time [4]. To address this methodological challenge and generate timely insights, digital health scholars have operationalized research protocols that capitalize on the unique characteristics of these technology mediated interventions [5-7]. Scholarship on optimizing digital health trials has ranged from leveraging smartphone ubiquity to accelerate research recruitment and informed consent [8], to harnessing sensor capabilities to capture novel clinical endpoints [8], to employing research designs and methods from the engineering sciences that can rapidly yield actionable outcomes [9].

Recently, the application of *digital health analytics*, defined as "the discovery and communication of patterns in health data," has emerged as an efficient approach to evaluate digital health interventions [10]. Numerous evaluative endeavors have mined the rich log data generated by users engaging with these inventions and successfully generated evidence of their impact on health outcomes [11]. In turn, the analytic models derived from these efforts have enhanced intervention effects through identifying the mediating behavioral mechanisms that motivate improved outcomes [12-14]. Scholars have also pushed for new theoretical frameworks to guide more systematic log data analyses and support transparent and replicable evidence generation [15]. The aggregation of this research productivity has advanced the scale and spread of research analytics in industry and academic evaluative practice.

In March 2018, our research group sought to contribute a resource to support applying analytic research methods to evaluate digital health interventions. The Analytics Platform to Evaluate Effective Engagement (APEEE) with digital health interventions was developed to facilitate the quantification, analysis, and visualization of research data [16]. The platform provides investigators with the means to characterize the breadth and depth of digital health engagement required to change behaviors and achieve intended health outcomes. With APEEE, investigators are able to cull through large, dense, and dynamic datasets in real time and identify meaningful patterns of digitally mediated behavior change. They can apply this functionality to conduct analytic evaluations and generate timely evidence to optimize intervention effectiveness. A formative evaluation of APEEE showed that digital health researchers perceived the platform to be an acceptable evaluative resource and were satisfied with its design, functionality, and performance [16]. They saw potential in APEEE to accelerate and augment evidence generation and expressed enthusiasm for adopting the platform to support their evaluative practice once fully implemented. However, more implementation research was required to formally evaluate the impact of the platform on digital health evaluative practice.

Although analytic methodologies have shown promise as an alternative approach to evidencing digital health interventions, they have not been without limitations. Trials have been small, research processes are often ad hoc and reactive, and barriers to introducing and sustaining new practices are rarely discussed [17]. Traditional research operations may be slow and cumbersome, but they benefit from a legacy of checks and balances that have been honed to ensure the valid collection of outcome measures [2]. There is currently no equivalent to these standard operating procedures to address analytic issues of missing or erroneous trial data, owing to digital device failure or participant disengagement from these devices [18]. Concerns regarding privacy and data security in digital health care delivery have also extended to digital health research [19]. Aggregating data from multiple sources for the purpose of applying research analytics will require the adoption of standards, raise privacy and ethical concerns, and demand new ways to preserve privacy [20]. Significant time, effort, and resources are required to develop the capacity to conduct data-driven research, particularly to set up the technological infrastructure, and also to train and support both staff and patients in this new practice.

Objectives

In short, there is a growing mismatch between the promising evidence base emerging from digital health analytics and the complexity of introducing these novel research methods into evaluative practice. Limited research exists on implementation challenges, notably the policy and regulatory climate required to support new approaches to digital health evidence generation, the practicalities of organizational change to accommodate analytic methodologies, and the personal narratives of digital health innovators who engage in these practices. To address these knowledge gaps, we sought to explore a new implementation of research analytics in a digital health research lab at the Hospital for Sick Children (SickKids) in Toronto, Canada, and draw wider insights on the contextual factors that enable and constrain analytic models of evaluation.

Methods

Research Overview

This study aimed to generate transferable insights into the process of implementing research analytics to evaluate digital health interventions.

We sought to answer the following two research questions: (1) how should the service of research analytics be designed to optimize digital health evidence generation and (2) what are the challenges and opportunities to scale, spread, and sustain this service in evaluative practice?

Our objectives were as follows:

1. At the macro level, to review policy and regulatory mandates for evidencing digital health interventions and assess whether research analytics can generate acceptable levels of evidence.
2. At the meso level, to explore the sociotechnical systems and services that enable operationalizing research analytics into evaluative practice.
3. At the micro level, to understand the personal motivations of digital health innovators for applying research analytics to evidence their products.

Study Design

We conducted a qualitative multilevel embedded single case study of implementing digital health research analytics in evaluative practice that comprised a review of the policy and regulatory climate in Ontario (macro level), a field study of APEEE's introduction into evaluative practice (meso level), and interviews with digital health innovators on their perceptions of analytics and evaluation (micro level). Employing an embedded single case study design allowed us to coalesce multiple levels of mutually shaping contexts that influence a representative case of digital health evidence generation [21]. All embedded units of analysis are summarized in Table 1 and described below.

Table 1. Overview of embedded units of analysis.

Embedded units of analysis	Type and nature of data
Macro-level study of the provincial context	Provincial policy and regulatory documents; provincial stakeholder working group field notes and meeting minutes; informal unstructured interviews with 5 senior leaders
Meso-level study of organizational change	Accounts of 5 research staff involved in conducting digital health evaluations; approximately 20 hours of observations in a digital health research lab; implementation artifacts
Microlevel study of personal motivations	Formal semistructured interviews with 21 digital health innovators

Data Collection

Macro Level

A broad literature review of recent provincial documents (eg, policies, frameworks, and funding programs) was conducted to assess the current digital health context. We sought to discern strategic interests for evidencing digital health interventions, with a specific interest in identifying barriers and facilitators to implementing analytic research methods. In parallel, between January 2017 and March 2019, we attended 7 working group sessions on sustaining and evaluating digital health innovations hosted as part of *Project SPARK*, a collaborative partnership between the Ministry of Health and Long-Term Care (MOHLTC), eHealth Ontario, the MaRS Discovery District, and the University Health Network in Toronto, Canada [22,23]. This partnership was formed to bring together thought leaders who shared a common vision to stimulate consumer digital health innovation in Ontario. Approximately 10 to 15 stakeholders across partnership organizations were present at each meeting. Topics discussed included the challenge of sustaining and evaluating digital health applications that were connected to provincial health assets (eg, Ontario Laboratories Information System and Digital Health Drug Repository). Field notes and meeting minutes were retained for analysis. Finally, we conducted informal unstructured interviews with a convenience sample of 5 senior leaders from academic (n=3) and industry (n=2) organizations to seek corroboration and clarification on the policy and regulatory directives identified in our literature review. These interviews were not audiotaped in accordance with informant preferences, but we obtained verbal consent to record interview notes for analysis.

Meso Level

To gain a real-world understanding of how to introduce digital health research analytics into evaluative practice, we conducted a 6-month field study of APEEE'S inaugural implementation in the Improving Outcomes in Child Health through Technology (iOUCH) lab at SickKids. The iOUCH lab aims to improve the lives of children and adolescents through the use of innovative information and communication technologies [24]. The research group comprises a principal investigator, a PhD-prepared associate, numerous managers, coordinators, and analysts, and a rotating roster of students and fellows, for a total of over 20 research staff. The group conducts research to conceptualize, design, and evidence digital health interventions and outsources the development of the interventions to external research groups or software development studios. This field study follows the preliminary work conducted by our research group to design and develop APEEE and connect it to *iCanCope*, a smartphone-based pain self-management program tailored for young people aged 12 to 25 years with chronic pain [25]. The app prompts them to check in every day and report their symptoms, set goals to improve their pain and function, read about pain coping strategies, and get social support from other young people living with chronic pain. *iCanCope* was conceived by the iOUCH lab, with our research group as the development partner [26]. The app also served as a use case to inform APEEE's product specifications during the design and

development of the platform [16]. This existing partnership and history of research collaboration provided a primed setting for us to study the reality of setting up and delivering the service of research analytics in a digital health research lab.

We aimed to map the network of people, tasks, and organizational routines that were required to support APEEE and the changes in these interactions and interdependencies over time. This sociotechnical approach was operationalized through conducting 5 half-day observation sessions with 5 research staff (ie, 1 associate, 1 manager, 1 coordinator, and 2 analysts) who worked on the *iCanCope* project. We observed staff engaging in their daily routines throughout the course of the study to capture changes in evaluative practice. Field notes were recorded during each observation session for analysis. We supplemented these observation sessions through conducting 30-min interviews with all 5 staff members at the start and end of the field study. A semistructured interview script was used to elicit expectations for how APEEE might change practice and, consequently, whether these changes occurred as expected. In addition to the abovementioned research activities, we were in a unique position to be responsible for supporting the introduction of APEEE into the iOUCH lab and the subsequent maintenance of the service. In our role as the APEEE product team, we conducted training sessions, made ad hoc changes and updates to the platform, drafted guidance documentation, and provided ongoing client support. This environment allowed us to experience a vendor-client relationship and react to real implementation challenges. Through this study, we were able to study the process of implementing APEEE in ethnographic detail and catalog the services and artifacts that were produced, both planned and improvised.

Micro Level

We approached 33 digital health innovators to participate in a 30- to 45-min audiotaped semistructured interview on their perceptions of digital health evaluation and analytics. A purposive sampling method was employed where innovators who were known to the research team were invited to participate in this research. We also sought maximum variation in innovator sector, occupation, and rank to capture a broad range of contexts and motivations.

The following criteria were applied to purposefully select digital health innovators for inclusion in this study:

1. Adults aged 18 years or older
2. Conversant in English
3. Identified by the research team as a digital health innovator (we define a digital health innovator as both academic and industry professionals who are involved in the design, development, and evaluation of digital health interventions; this includes scientists, clinicians, project managers, research coordinators, chief executive officers, chief technical officers, statisticians, developers, designers, data architects, and other relevant positions)
4. Previous or current involvement on a digital health project
5. Interest in evaluating the impact of a digital health intervention

All identified innovators were first contacted through a study recruitment email sent by the lead researcher. This email contained a brief description of the study objectives as well as a succinct summary of study activities and considerations (eg, data privacy and confidential quotes from any transcripts). If innovators expressed interest in joining the study, they were scheduled for an interview. Consent was obtained verbally and audiotaped before the start of the interview. Participants were not compensated for their involvement in this research. Ethics approval for this study was obtained from the University of Toronto Research Ethics Board (reference number 00035682).

Data Analysis

Analytical Framework

Our analysis drew from Greenhalgh et al's diffusion of innovations model to characterize the complexity of innovating digital health evaluative practice [27]. This theoretical model has been widely used to study the scale and spread of health innovations into organizational practice. We were motivated by Greenhalgh et al's own use of this model in a case study exploring the introduction of shared electronic patient records (EPR) into care sites in the United Kingdom [28]. Although EPRs may be specific to health care settings, the data sharing mechanisms that power them are context-agnostic. The access that EPRs provide to clinical data networks is comparable with the access that analytic resources provide to research data repositories. As such, the diffusion of innovations model allowed us to explain and interpret (1) the complexity of implementing digital health research analytics in evaluative practice; (2) the dynamic nature of the implementation process; and (3) the shifting political, organizational, and technological context over time. We also heavily referenced Greenhalgh et al's work on the real-world implementation of video outpatient consultations in UK clinical practices to structure our research according to macro, meso, and micro levels of analysis and produce a cohesive story of how the digital health research system reacts to methodological change [29].

Process

Our efforts to map the sociotechnical digital health research system provided detailed data on the challenges of operationalizing research analytics and the workarounds to overcome them. We amassed data on issues related to the technology, the research environment, administrative processes, and data privacy and governance. This research generated various types of qualitative data (ie, documents, field notes, and interview transcripts), which we analyzed separately using the directed content analysis approach [30]. With Greenhalgh et al's diffusion of innovations model as our conceptual anchor, we identified key components from the model to serve as initial coding categories [27]. We developed operational definitions for all relevant components of the model, assigned a code to each component, and then sorted through our data to identify themes and phrases that could be organized under our theory-informed codes. We used first-order codes to categorize our data (eg, *attributes of the technology as an innovation* and *organizational antecedents for innovation*), with second-order codes fleshing these out in greater detail (eg, under *organizational antecedents for innovation* were subcategories

of *relative advantage* and *organizational slack*, among others). Once coding was completed, we thematically analyzed within and across the three embedded units of analysis to detail the reconfiguration of organizational routines that accompanied the introduction of research analytics. This thematic investigation allowed us to surface the tensions between current *research as usual* and new ways of conducting digital health research facilitated by analytics. Our findings are reported in a narrative case format [31,32], with direct quotes and attributions embedded into the text (eg, P1) to convey the style and intent through which participants expressed their thoughts and experiences [33].

Results

Macro-Level Findings

Currently, there is a strong policy push in Ontario to build digital health capacity. This agenda was initiated in February 2015 by the MOHLTC with the launch of their *Patients First: Action Plan for Health Care* [34]. Designed to "put people and patients first by improving their health care experience and their health outcomes," the plan outlined four key objectives to place the patient at the center of the health care system and shift care from hospital to home: (1) provide faster access to care, (2) deliver connected care in the community, (3) keep patients informed to facilitate their health decisions, and (4) protect the public health care system through sustainable policy practices. In support of this proposed health system transformation, the *Patients First: Digital Health Strategy* was published in November 2016 to "advance modern, integrated, patient-centred care" [35]. During his keynote presentation at the 2016 Canada Health Infoway Partnership Conference on the strategy, Deputy Minister Robert Bell reflected that achieving public *Patients First* commitments would require "three building blocks of digital health: strategy, governance, and information management" [35]. This foundation was conjectured to enhance access to health information and services, strengthen health care quality, and stimulate economic growth.

Central to these tenets was the belief that digital health represented positive innovation to the health care system and thus warranted building out infrastructure to modernize existing practices. Of the 7 "guiding principles for digital health" laid out in the digital health strategy, the *Digital First* philosophy was notable in its recommendation that all new and existing programs should be assessed by asking, "how can we do it with digital health?" This enthusiasm was echoed in two major reports commissioned by the Government of Ontario to assess the value of its digital health assets: the Ontario Health Innovation Council and the Advisory Council on Government Assets submitted reports that lauded the potential for the digitalization of Ontario's health care system to generate "significant and ongoing value and opportunities for patients and families, providers, and the economy" [36,37]. These reports were widely endorsed by industry, academic, and professional organizations across Ontario [38-40]. They consequently led to the creation of numerous strategic digital health funding programs, notably the 4-year Can $20-million Health Technologies Fund to support the early evaluation, procurement, and adoption of innovative technologies into the provincial health care system. In January 2017, the MOHLTC developed the *10-point Digital Health Action Plan* to operationalize its Digital Health Strategy [41]. Both consumer-facing and health system–facing initiatives were identified, and a Digital Health Scorecard was introduced to measure success against quantitative key performance indicators, some of which were ambitious in scope. For example, the scorecard mandated that the number of patients who used a digital health intervention annually should go from 4600 in 2017 to 100,000 in 2021. It was unclear how baseline and projected figures were estimated, what measurement mechanism would capture this growth, or what would happen if performance indicators were not met.

We did not find any provincial policies specifically related to evidencing digital health interventions. The paucity of guidelines or frameworks on acceptable evidence for digital health care technologies made it difficult for us to discern whether analytic research methods were up to standard. Although most of the policy documents we reviewed referred to digital health innovations as being "effective," "efficient," "beneficial," and poised to deliver "value…in areas such as health outcomes, cost avoidance, and jobs" [37], there was no mention of the methods used to evaluate these attributes. However, this hype did not go unquestioned, as concerns were raised by provincial-level stakeholders and documents regarding the effect of health technologies on patient safety and quality of care [36]. Informants noted that there were no systems in place to measure and evaluate the value of these innovations on health outcomes, cost avoidance, and job creation. These apprehensions were primarily assuaged through assurances that "a culture of innovation" led by strong leadership across all levels of the health care system and supported by effective change management would succeed in promoting the adoption of "beneficial health technologies" [42].

Indeed, policy makers saw greater value in brokering partnerships between technology companies and evaluation experts to evidence digital health technologies, in lieu of releasing evidence standards to support building internal evaluation capacity. This belief was operationalized through the launch of two provincial funding programs dedicated to the evaluation of digital health technologies. From 2016 to 2018, 26 digital health technologies were funded a total of Can $10.4 million through the Health Technologies Fund [43]. To qualify for funding, applicants were required to assemble a team comprising a public health service provider organization, a community-based association or advocacy group, a for-profit health technology business, and an established evaluation provider. Teams were not required to use a particular research design or meet a predetermined threshold of evidence. Rather, evaluating a digital health technology served as a strategic means to a specific end goal: the procurement of this technology into the Ontario health care system. In contrast, the province's decision in October 2017 to allocate Can $1 million in funding for a new Center of Excellence (COE) in Digital Health Benefits Evaluation reflected the renewed value of evidence as a stand-alone policy lever [44,45]. The COE's mandate was to form a consortium of evaluation partners who would be responsible for conducting the majority of the province's digital

health evaluations and generating "responsive, timely, rapid, robust, high quality evaluations of both innovative and mature digital health assets and digital health technologies" [44]. This call marked the first instance of the province specifying the need for evaluations that could keep pace with the digital health technologies under study. Support for both these initiatives has been mixed; provincial stakeholders endorsed their potential to standardize approaches to digital health evaluation, whereas industry informants expressed reservations that few innovators would manage to secure the partnerships required to qualify for this increasingly exclusive evaluation expertise.

Although provincial policy makers have taken a permissive stance on digital health evidence generation, the federal regulatory bodies that govern these technologies have been more prescriptive. In January 2019, Health Canada released a draft guidance document on the regulation of Software as a Medical Device [46], which sought to classify medical software as meeting the definition of a medical device under the Food and Drugs Act [47], thereby requiring compliance to the Medical Device Regulations [48]. Health Canada also issued a notice of intent to strengthen the postmarket surveillance risk management of class II to IV medical devices [49] and cited the increased availability of real-world data and evidence generated by devices as rationale for this change [50]. The agency intends to propose a series of significant changes to the Medical Device Regulations by Fall 2020, notably providing the Minister of Health with the authority to request (1) *analytical issue reports* from a medical device manufacturer on suspicion of a safety concern and (2) *annual reports* on medical device performance on safety and effectiveness targets [49]. The adoption of analytics as a mechanism to enforce regulatory compliance is significant in its recognition of log data as a valid measure to inform product safety and effectiveness. This enhanced use of real-world evidence throughout the product life cycle was well received by provincial-level stakeholders, who perceive this approach to be aligned with the Digital Health Strategy's mandate for "increased transparency to guide governance decisions" [35]. Our industry informants were less enthused about potentially having to disclose proprietary information regarding their product's market performance. They also expressed concerns that failing to meet regulatory targets would have downstream effects on their eligibility for government funding if analytic reports were shared across agencies.

Overall, our analysis identified a small number of policy and regulatory programs with a minor focus on digital health evidence generation, supported through pockets of provincial funding. Provincial-level stakeholders and documents agree in principle that digital health interventions should be supported by a robust body of evidence to warrant public funding. However, the lack of a definitive guidance on what constitutes *good evidence* puts onus on digital health innovators to self-assess the evaluative approach that will yield the greatest return on investment, or to apply for provincial funding schemes that will pair them with an academic evaluation partner. This emerging shift toward a small number of academic groups conducting a large proportion of the province's digital health evaluations presents an opportunity for the academic sector to drive the provincial digital health research agenda. Although the adoption of research analytics by this sector would rapidly scale this methodology in evaluative practice, the need for sustained funding to maintain the data infrastructure required to operationalize analytic methods is likely to be a significant barrier to provincial rollout.

Meso-Level Findings

The inaugural implementation of APEEE in the iOUCH lab was characterized by periods of organizational acclimation with intermittent technological iteration. Despite the fact that the research group had been previously engaged in the design and development of the platform and informed its inaugural connection to *iCanCope* [16], implementation proved far more complex and challenging than anticipated. Our approach to introducing the platform into evaluative practice was highly coordinated: we established a small team to manage operations, conducted (1) one 2-hour on-site training session with 3 research staff and (2) two 1-hour video training sessions with 2 research staff to instruct them on APEEE features and functionality, opened a support channel on the Slack collaboration tool to facilitate ongoing communication, and provided staff with guidance documents (eg, product manual, data dictionary, and *frequently asked questions* content) to reference throughout the 6-month trial period. From October 2018 to March 2019, 5 members of the iOUCH research group were able to access APEEE through individual accounts and execute real-time queries across 2.5 years' worth of *iCanCope* engagement data.

Before APEEE's introduction into the iOUCH lab, there were signs that the research group was interested in exploring alternative methods of digital health evaluation. Much of the research conducted on *iCanCope* used the randomized controlled trial (RCT) design to generate evidence of clinical efficacy [51]. In their prestudy interviews, research staff disclosed that while they recognized the importance of definitive trials, they also saw value in conducting "quick tests" (P4) to optimize app performance and improve participant engagement. As such, staff were keen to trial novel research methods that would enable them to move away from "business as usual" (P14) and address research questions that their RCTs were not designed to answer. APEEE's compatibility with these emerging organizational values lowered barriers to adoption but also raised expectations for the platform's capacity to transform existing practices.

Throughout the course of this study, research staff maintained a perception of APEEE as a "useful" (P1), "reliable" (P2), and "effective" (P13) research resource that "did everything it was supposed to do" (P9). One staff member remarked in their exit interview that the platform "is fantastic but also expected—how could you do this kind of research and not have something like this?" (P9) This positive appraisal of the platform was partly because of its relative advantage in comparison with the analytic process that it ultimately replaced. Before APEEE's implementation in the iOUCH lab, research staff had to run a command in the Mac operating system Terminal emulator and extract a large comma-separated values file of tabular log data. They were then forced to sift through thousands of log records and parse out the appropriate data fields to resolve their query. With the exception of one staff member who was comfortable building pivot tables in Excel to expedite analyses, all other

staff members found the process to be "so overwhelming" (P14) and "unsustainable" (P13) given the increasing volume of participants to be enrolled into trials of *iCanCope*. Overall, staff lamented the "tedious" (P2) and "time-consuming" (P1) effort required to answer "basic questions" (P13), and obtain a "quick snapshot" (P2) of study status and performance. Thus, they saw APEEE as a welcome change to existing processes and stated at their exit interview that they would recommend "anyone doing research on apps like *iCanCope* to have access to APEEE" (P9).

One key attribute of APEEE that eased its acceptance and adoption by research staff was the degree to which the platform could be customized according to staff roles. As part of the service of delivering APEEE, we embedded a needs assessment into our prestudy interview with staff. We were consequently able to build out custom dashboards and visualizations with these needs in mind. Figure 1 presents an APEEE dashboard designed for a research assistant whose primary task is to make weekly calls to study participants and encourage them to engage with *iCanCope*. This dashboard has focused content: only 4 analytic indicators are displayed, and the sidebar has minimal features. Figure 2 presents an APEEE dashboard for a research coordinator who manages multiple streams of research activity, from monitoring participant recruitment and adherence to the study protocol to conducting statistical analyses across study outcomes. In contrast, this dashboard covers a breadth of analytic indicators (only 10 are visible but 12 additional indicators can be viewed through scrolling down the dashboard), allows access to more features, and also has data export functionality to support offline work and interoperability with other analytic resources. Staff responded positively to the adaptable nature of the platform and took advantage of this capacity for customization through making numerous requests to "tweak" (P9) existing dashboards or build new ones to support different research tasks.

Figure 1. Screenshot of the Analytics Platform to Evaluate Effective Engagement research assistant dashboard.

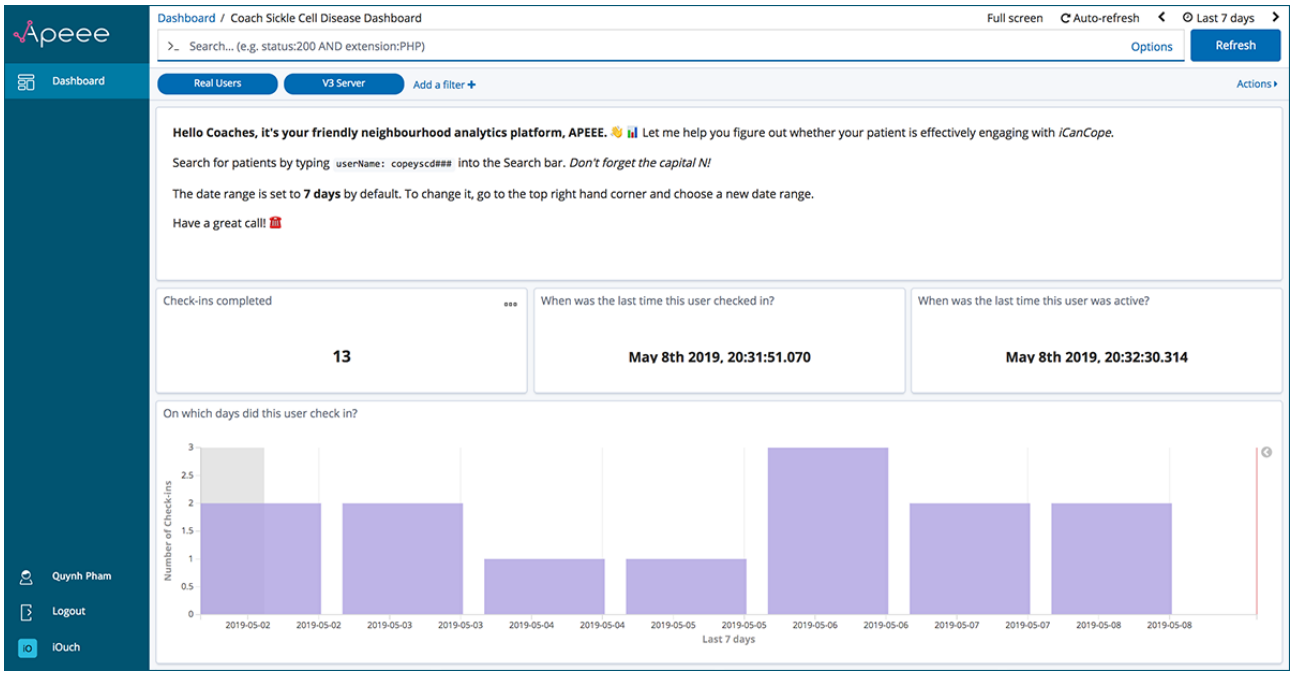
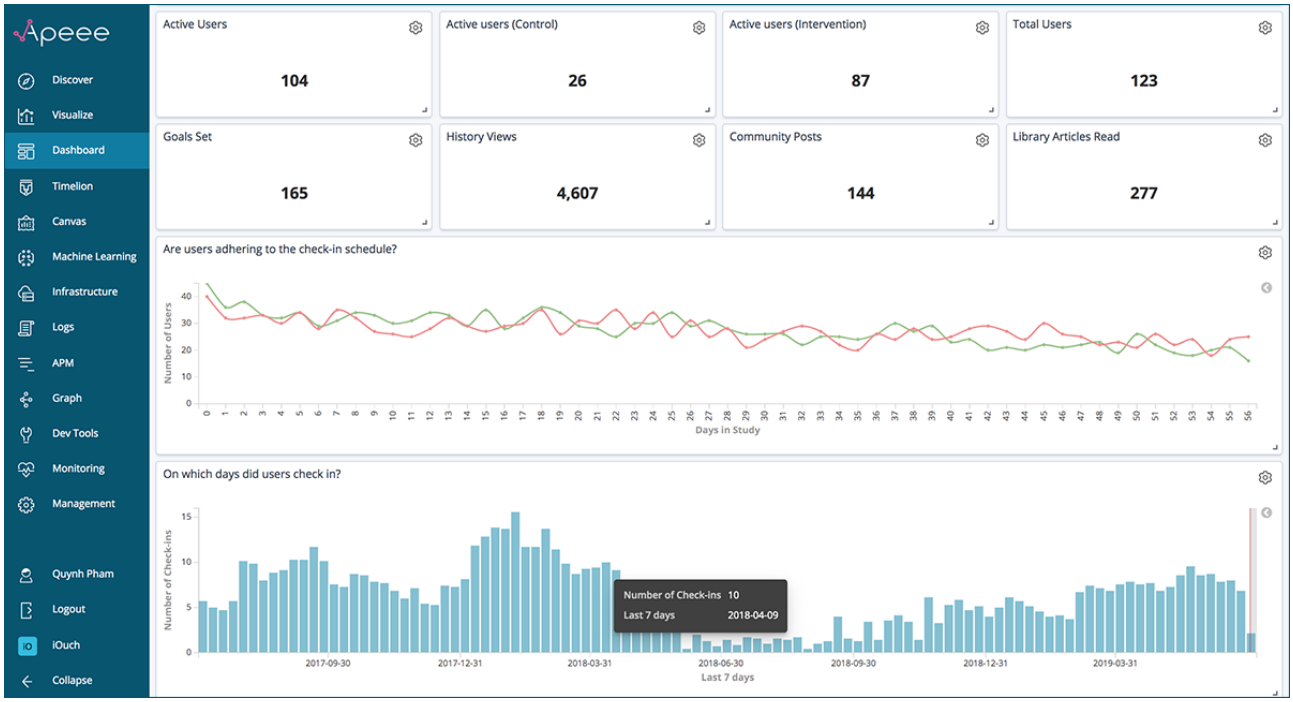


Figure 2. Screenshot of the Analytics Platform to Evaluate Effective Engagement research coordinator dashboard.



Although these requests were initially perceived by our research team as a sign of implementation success, it was only after fulfilling them that we noticed iOUCH research staff were asking for visualizations that they did not actually use. During our on-site observation sessions, staff almost always accessed the same two analytic indicators: (1) a data table of participants who had checked into *iCanCope* alongside the number of check-ins they had completed and the date of completion, and (2) a metric of the number of active users, defined as users who had generated any log data. This behavior was highlighted toward the end of the field study when our internal server unexpectedly reached a log data storage limit, forcing our research team to create *lite* versions of all APEEE dashboards with a subset of analytic indicators to maintain performance. When asked in their poststudy interview whether only having access to a limited number of indicators had affected the utility of the platform, staff commented that they "hadn't even noticed a difference" (P2). The majority of indicators that had been made available to them were "interesting" (P13) and "nice to have" (P1), but ultimately did not change or improve their daily practices. However, they elaborated that had APEEE been taken down completely, this would have significantly impeded the data extraction and analysis routines that had been established around the platform.

Several characteristics of the iOUCH lab and its evaluative practice may have served as organizational antecedents for research innovation. APEEE's introduction into the lab was championed by senior investigators who directed the research agenda and had good managerial relations with research staff. They were able to convince staff that the platform—and research analytics more broadly—was synergistic with existing projects and would facilitate research processes. In terms of staff competence, we noted a high *absorptive capacity for new knowledge*, defined by Greenhalgh et al as "a combination of formal expertise, informal organizational know-how, technical infrastructure, and relevant interpersonal networks" [28]. Finally, there was sufficient *organizational slack* within the lab, defined by Greenhalgh et al as "spare time, money, or expertise that can be channeled into new projects" [28], for staff to bear the onboarding process and develop confidence in using APEEE.

Unfortunately, this organizational slack did not extend to *methodological slack*, which we define as the capacity to pursue study designs and methods that significantly deviate from current research practice. Despite high interest and intention for APEEE to change evaluative practice, the iOUCH lab was ultimately bound by the commitments they had made to evidence their apps using RCTs. We observed the group discovering through APEEE that users were not adhering to the study protocol (ie, not using the app to report their symptoms once a day, every day, for 56 days) and also failing to use certain app features and functionality. However, staff had no bandwidth to react to these analytic insights, given the methodological rigidity of the RCT design to maintain internal validity [52]. This tension played out in surprising ways, the most notable being that staff eventually realized the limitations of applying APEEE to effect methodological change within an RCT-locked environment, yet still spoke highly of future opportunities beyond the field study to conduct data-driven trials. It is worth noting that toward the end of the field study, the lab had started drafting a secondary analysis of their RCT data comprising subgroups identified through APEEE and more broadly exploring alternative research designs (eg, sequential multiple assignment randomized trials and real-world observational studies) to maximize platform utility [53].

A critical factor that both enabled and constrained APEEE's capacity to effect methodological change in the iOUCH lab was the underlying *iCanCope* data model. Before APEEE's connection to *iCanCope*, there was no working hypothesis or data analysis plan to characterize the relationship between engagement and health outcomes. As a result, the set of analytic

tags built out to log events and generate data for ingestion into APEEE did not capture the full breadth and depth of interactions that could have denoted effective engagement with the app, defined as "sufficient engagement to achieve intended outcomes" [54]. Furthermore, the structure and nomenclature of the tags themselves made it difficult to visualize the desired analytic indicators on APEEE, given the specific requirements of the platform. This architectural incompatibility allowed us to grasp the interdependencies between the data going into APEEE and the insights coming out. From this, we identified the need to expand the service of APEEE to include consultations with groups on (1) the research questions they want the platform to help them answer, (2) the parameters of their current data model and analytic tags to answer select questions, and (3) the amendments or additions to their data infrastructure required to answer all questions—to be done by their development team or our own.

Our findings suggest that the capacity for research analytics to optimize digital health evaluations is highest when there is (1) a collaborative working relationship between research client and analytics service provider, (2) a data-driven research agenda, (3) a robust data infrastructure with clear documentation of analytic tags, (4) in-house software development expertise, and (5) a collective tolerance for methodological change. Although these success factors are not listed in any particular order, we wish to emphasize the significance of establishing trust when introducing research analytics into evaluative practice. When asked to specify facilitators of APEEE's introduction into the iOUCH lab, research staff unanimously cited the support provided by our research team as a determinant of success. They espoused the *process* of implementing APEEE—meeting to review the *iCanCope* data model and select appropriate analytic indicators, codesigning and iterating on dashboard design and content, engaging in meaningful dialog about analytic insights—as being more valuable than the *product* itself. This response supports our intent to position the platform as the technical *hard core* of a proposed research analytics service, with a suite of adaptable evaluation services forming the *soft periphery* [55,56]. Multimedia Appendix 1 presents our preliminary service blueprint of this research analytics service, which draws together the human and technical interactions and interdependencies required to implement this service in evaluative practice.

Microlevel Findings

A total of 21 digital health innovators agreed to be interviewed for this research; reasons for denying the request included lack of availability (n=9) and professed lack of knowledge on analytics and evaluation (n=3). On average, innovators were young, male, educated, and worked in academia on mobile-based digital health interventions. Table 2 details their demographic characteristics, and Multimedia Appendix 2 presents the semistructured interview script.

In all the interviews we conducted, regardless of sector, role, or rank, innovators were able to clearly communicate the problem or unmet need that their product aimed to address. They provided empathetic and nuanced descriptions of how chronic diseases "totally change the way people go about their everyday lives" (P7) and were able to explain complex clinical concepts (eg, prostate-specific antigen nadir, and etiology of heart failure) without formal medical training. Innovators were "excited" (P10), "proud" (P12), and "satisfied" (P17) to work on projects that "made a real difference" (P6) and were "on the cutting edge of health care" (P8). When asked to describe how their product aimed to solve the problem they had identified, most innovators referenced a conceptual or logic model that delineated the solution. These innovators also provided measured estimates of their product's potential effect on health outcomes, often adding caveats that more "research" (P19), "users" (P16), or "time" (P3) was required to validate claims. In contrast, a small number of innovators found it difficult to convey the mechanisms of change that their product would facilitate to improve health outcomes. They also held ambitious beliefs of their product's capacity to positively impact health and well-being. When probed to substantiate the rationale supporting these beliefs, innovators were hesitant to elaborate and defaulted to reiterating product features and functionality.

Overall, innovators were able to communicate the ideal or intended user journey for their products with ease and were confident in their assessment of what constituted effective engagement with their product. All innovators defined effective engagement quantitatively (ie, amount, duration, breadth, or depth of intervention usage); some identified a single event that was critical to users deriving any benefit, whereas others listed a series of sequential events. We noted a shared prioritization on events that involved capturing data, for example, entering blood glucose readings or logging asthma exacerbations. When asked whether users were effectively engaging with products as intended, half of all innovators disclosed that they were "not sure" or "needed to check." The other half said yes; however, only 5 innovators referenced formative research or definitive trials to support their claims.

Table 2. Demographic characteristics of digital health innovators.

Characteristics	Values
Age (years), range	22-45
Age (years), mean (SD)	30.4 (5.3)
Age (years), median (IQR)	29 (4)
Gender, n (%)	
Male	13 (62)
Female	8 (38)
Education, n (%)	
Graduate	17 (81)
Undergraduate	4 (19)
Sector, n (%)	
Academic	18 (86)
Industry	3 (14)
Occupation, n (%)	
Research coordinator	4 (19)
Research analyst	3 (14)
Designer	3 (14)
Developer	3 (14)
Manager	3 (14)
Director	2 (10)
Product owner	2 (10)
Research associate	1 (5)
Innovation platform, n (%)	
Mobile	13 (62)
Web	8 (38)
Clinical focus of innovation, n (%)	
Well-being	4 (19)
Heart failure	4 (19)
Chronic pain	3 (14)
Juvenile idiopathic arthritis	2 (10)
Mental health	2 (1)
Prostate cancer	2 (10)
Sickle cell disease	2 (10)
Asthma	1 (5)
Diabetes	1 (5)

Of the 21 innovators interviewed for this research, every single one included the word *data* in their definition of analytics. Innovators broadly defined the practice of analytics as the collection, aggregation, or visualization of data to generate actionable insights. They sought to differentiate “raw data” (P11) from analytics, which they saw as “data with meaning” (P7) that could be used to “understand the current state of a product and predict how people might use it in the future” (P15). They also endorsed the notion that “data never lies” (P5) and consequently perceived analytics to be a valid source of information to “help with big decisions” (P17). Innovators intuitively considered analytics to be a quality improvement or project management initiative and did not mention its application in evaluative practice. They valued the opportunity provided by analytics to “sanity check design and development assumptions” (P18) and “get to know users better” (P4), with a final aim of translating this knowledge into product specifications. They also emphasized an expectation that analytic insights be generated in real time and visualized in a way “that didn’t take much time to read through and understand” (P11).

Almost all innovators referenced the Google Analytics platform when describing their familiarity with analytics in practice; some even initiated their definition of analytics with mention of the platform. Interestingly, of those 19 innovators, only 4 had ever personally used the platform and found it to have a "steep learning curve" (P4), "really confusing to use" (P10), and "totally impossible to segment users into useful subgroups for analysis" (P21).

Despite these challenging experiences with Google Analytics and an overall lack of experience with analytics across the group, innovators unanimously expressed interest in applying research analytics to evidence their products. They also shared similar preferences for how analytic insights should be presented, namely through a dashboard interface with dynamic widgets containing aggregate-level analytic indicators. However, there were striking sectoral differences in the quality of evidence innovators believed research analytics could generate, as well as its intended use. Industry innovators espoused the opportunity for analytics-enabled research to "replace trials" (P20) and "generate real-world evidence" (P18) to support procurement efforts and improve market valuation. They disclosed that they were already collecting usage logs and patient-reported outcomes but lacked a resource to collate these data and interpret insights. In contrast, academic innovators saw value in conducting research analytics to inform subgroup analyses for a definitive trial or to generate quantitative data to complement qualitative research. As a stand-alone source of evidence, there was consensus among academic innovators that analytic insights placed "near the bottom of the evidence pyramid" (P2) and were insufficient on their own to demonstrate efficacy [57]. However, academic innovators saw value in research analytics as a part of a broader "data-driven research agenda" (P1), and affirmed that applying analytic insights to optimize digital health interventions might help them to "survive definitive trials and demonstrate efficacy" (P9). Other areas of application proposed by the group included (1) monitoring the status of multisite studies to ensure standardized study protocol execution; (2) conducting A/B testing on product features and functionality to "test out behavioural hunches" (P15); (3) identifying "dying or dead" (P19) intervention components that had limited effects on outcomes; and (4) informing strategies to promote adherence and prevent disengagement from both the study and the product.

Although innovators were encouraged by the proposed benefits of research analytics, they also acknowledged numerous barriers to using analytic insights in practice. Executive-level innovators divulged that they often made decisions based on "intuition" (P21) or "gut feeling" (P15), and were unsure of how they would react to data that conflicted with their convictions. All three developers raised issues of personal health information being openly used in a nonclinical context and cautioned that consent forms and terms of agreement would need to be updated to reflect this "off-label use of log data" (P5). Across the group, the most cited challenge to adopting research analytics was a collective reservation that analytic insights could be "trusted." Some innovators voiced concerns that users often behaved in "random and unpredictable ways" (P10), and thus, the data they generated might not be reflective of their self-care intentions. Others indicated that they would need to be certain of log data quality, specifically "data accuracy and comprehensiveness" (P9), before using analytic insights to inform decision making.

Discussion

Principal Findings

This study illustrates the complexity of implementing research analytics in evaluative practice to evidence a digital health intervention, taking into account the political, organizational, and personal contexts that influence methodological change. Our findings confirm that the practice of research analytics is an efficient and effective means of supporting digital health evidence generation. The introduction of an APEEE with digital health interventions into a busy research lab was ultimately accepted by the research staff, routinized in their evaluative practice, and optimized their existing mechanisms of log data analysis and interpretation. By the end of the 6-month field study, the research group had arranged to sustain this innovation past the field study period to support future digital health evaluations. Although these findings suggest that research analytics may be integral to modernizing digital health research models, the process of effecting methodological change was not trivial. Despite emerging policy and regulatory interest in digital health evaluation and a primed research organization with engaged staff, establishing a new model of digital health evaluation was challenging. The difficulties of changing evaluative practice to accommodate analytic methodologies were largely attributable to (1) a discrepancy between perceived analytic wants and actual analytic needs when conducting digital health evaluations, (2) a nascent data infrastructure that was not architected to generate robust analytic insights, and (3) a lack of methodological slack to trial data-driven study designs and methods.

The emergence of methodological slack as a distinct system antecedent for digital health research innovation is significant in its recognition of the tangible constraints imposed by positivist research paradigms and traditional models of scientific inquiry [58]. Although digital health innovations are increasingly recognized as complex interventions that deliver care within complex health systems [59], conventional approaches to evidencing them remain wedded to definitive trials that assume linear causality and control for complexity [60]. This theoretical orientation comes at a cost, namely, the incapacity to adopt data-driven research designs and methods that are dynamic in process and assume emergent causality. We posit that a lack of methodological slack is similar to the phenomenon of *paradigm paralysis*, which is defined as "the inability or refusal to shift worldviews and see beyond current theoretical models of thinking" and acknowledged as a "block to creativity, innovation, and change" [61]. From our exploration of macro- and microlevel contexts that influence digital health evidence generation, we offer two factors that may be perpetuating this positivist bias: (1) the role of federal funding agencies in directing the provincial digital health research agenda and (2) the assessment of log data and analytic insights as a weaker form of scientific evidence.

Although there are sparse pockets of provincial funding earmarked for digital health evidence generation, the majority

of this research is federally funded through the Canadian Institutes of Health Research (CIHR). As the leading federal agency responsible for funding health and medical research in Canada, CIHR directs the provincial digital health research agenda through releasing strategic funding calls that target key areas of research. Funding calls are often prescriptive on methodological approach, for example, a recent call for "applications focused on the development, integration and evaluation of electronic health innovations…to optimize the outcomes of patients experiencing transitions in care" required the use of a "pragmatic randomized clinical trial methodology" and provided a list of suggested trial designs [62]. These directives illustrate a significant criticism of CIHR: that it is "positivist in orientation" and designed to fund "short-term, hypothesis-driven evaluative studies" [63]. As a consequence, research groups such as iOUCH are incentivized to conduct definitive trials, even if such trials may not suit the immaturity of the intervention and are less likely to demonstrate success.

In recent years, CIHR has developed its mandate to be more inclusive of alternative research paradigms. The agency acknowledges that "publicly-funded research is problem-driven, and the nature of science itself has changed with the move from positivist linear explanations to complex systems-based research and explanatory models" [64]. To operationalize this mandate and "change the paradigm of how research is rewarded," CIHR launched the Innovative Clinical Trials (iCT) Initiative in 2016 with an annual budget of Can $11.7 million to "support the development and adoption of innovative and cost-effective trial methodologies" that are "alternative to traditional RCTs" and aim to "reduce the cost of conducting trials, reduce the amount of time needed to answer research questions, increase the relevance of research findings to patients, health care providers and policy makers…and maximize the use of existing knowledge and data" [65]. Examples of iCT designs provided by CIHR include adaptive, n-of-1, registry, and observational trials [65]. A total of 3 years on from the first iCT funding call, evidence of the initiative's impact on shifting research programs to include innovative research methodologies was demonstrated during our research timeframe: in March 2019, the iOUCH lab was awarded an iCT Catalyst Grant to conduct a multiple baseline study of *iCanCope*, which will include the use of APEEE to monitor weekly outcomes data and establish stable baselines [53]. Moreover, of the 20 catalyst grants that have been awarded since 2016, 4 have funded innovative trials of digital health interventions and 2 have specifically funded data-driven optimization trials, which we contend are the most ideal study designs to support through research analytics [66]. This progress suggests that strategic federal funding initiatives such as iCT are an effective policy lever for promoting acceptance and application of analytics-enabled research to optimize digital health evidence generation.

Our interviews with digital health innovators revealed a reluctance to adopt research analytics as a method of evidence generation owing to a lack of trust in the log data quality powering analytic insights. This unease was more pronounced in academic innovators than their industry counterparts and manifested itself in the limited degree to which this sector was willing to consider analytic insights as a viable outcome measure to demonstrate efficacy. This finding builds on a burgeoning body of literature highlighting the use of flawed, uncertain, proximate, and sparse (FUPS) data to inform research and care [67]. From our findings, we posit that log data generated from users engaging with digital health interventions fit the criteria for FUPS data. Engagement log data may be flawed, due to missing data or erroneously logged events; uncertain, due to differences in how data are behaviorally conceptualized along an engagement continuum; proximate, in that analytic indicators of engagement indicate that users may be engaging effectively with a digital health intervention but do not definitively confirm a relationship between engagement and intended outcomes; and sparse, in that a low volume of events within key subgroups due to disengagement and attrition may limit the possibility of statistical inference. In their case study of a large-scale FUPS dataset of child mental health outcomes following contact with specialist mental health services in the United Kingdom, Wolpert and Rutter opine that clinical researchers are heavily influenced by the paradigm of evidence-based medicine and trained to interrogate data by its ranking on the evidence hierarchy [68]. As a result, they may be predisposed to criticize or dismiss FUPS data, particularly if such data challenge strongly held convictions or interests. The authors further note that FUPS data are often used in charged and contested contexts where conclusions drawn from them have significant implications; careful thought must be given to how these data are weighted for decision making.

In the context of digital health evidence generation, we acknowledge that the practice of research analytics may be compromised by the FUPS nature of digital health log data. However, we also acknowledge the harsh realities of the current digital health research landscape, namely the paucity of evidence-based digital health interventions currently available to consumers and the "data-poor" approaches to evidencing them that cannot keep pace with technologic change [4]. At a time of unsustainable growth in the burden and cost of chronic disease management [69], the promise of digital health to transform care has not been delivered. We contend that a new standard of digital health evidence that is proportionate to the risk of the intervention and the magnitude of the decision being made will allow for the consideration of log data as a valid data source to optimize interventions and improve decisions. To safeguard against drawing misleading conclusions, we recommend that digital health innovators adhere to three key principles proposed by Wolpert and Rutter for analyzing FUPS data: (1) be honest and upfront about the limitations of log data to produce causal inferences, (2) be transparent in the statistical analyses used to derive analytic indicators, and (3) triangulate analytic insights with other information [68]. In doing so, these data can support meaningful dialogue between key stakeholders, including policy makers, regulators, innovators, and users, in relation to the impact of digital health innovations on health and care.

Strengths and Limitations

Through this research, we were able to comprehensively study and document the implementation of research analytics in digital health evaluative practice. Working with front-line and senior research staff, we developed standard operating procedures,

technical guidance documents, and research protocols for setting up and running a research analytics practice. We collected rich qualitative data across macro, meso, and micro contexts to characterize the complexity of designing and implementing the service of research analytics, and identified the barriers and facilitators to scale, spread, and sustain this service in Ontario. We were able to operationalize this service and gain detailed insights into organizational routines and how these changed with the provision of APEEE. A trial phase of APEEE continues within the iOUCH lab, and further work is ongoing to extend the service to other research groups and evaluation settings.

The findings of our research must be viewed in light of their limitations. First, we had difficulty recruiting innovators from industry into this study owing to their demanding schedules and a lack of obvious return on investment; as such, industry narratives may be underrepresented. We engaged in nonprobability purposive sampling to recruit digital health innovators for interviews, which may have biased our sample selection and limited the generalizability of research findings. To mitigate this bias, we expanded our sampling frame to include innovators who could be interviewed virtually or on-site at international academic conferences. Second, our research focused on the implementation of research analytics in a single digital health research lab. Although we assert that the iOUCH lab is representative of a typical evaluation setting, it is possible that a different lab with no prior relation to our research group may have found the implementation of APEEE more difficult and less beneficial. Finally, the capacity to generalize findings from this research was a trade-off that we considered carefully when selecting the case study methodology to direct this research. Unlike quantitative study designs, the goal of case studies is to produce *analytic generalizations*, defined by Yin as "the degree to which findings bear upon a particular theory, theoretical construct or theoretical sequence of events" [70]. Analytic generalizations are distinct from statistical generalizations in that they do not draw inferences from data to a population. Instead, analytic generalizations compare the results of a case study with a previously developed theory and seek to generalize theoretical insights as opposed to actual study results. In our case study, the emergence of methodological slack was a form of analytical generalization that drew from Greenhalgh et al's theoretical construct of organizational slack [27] and may be applied to future cases of implementing research analytics in digital health evaluative practice. The majority of insights from our three tiers of analysis (eg, our suggestions on the capacity for research analytics to optimize digital health evaluations) can and should be characterized as analytic generalizations.

Conclusions

Scientific methods and practices that can facilitate the agile trials needed to iterate and improve digital health interventions warrant continued implementation. As outlined in this paper, the service of research analytics may help to accelerate the pace of digital health evidence generation and build a data-rich research infrastructure that enables continuous learning and evaluation. Valid concerns exist regarding the formation of unfounded or opportunistic causal inferences based on flawed analytic insights. However, continuing to pursue evaluative practices that fail to raise the standard of digital health safety and effectiveness is untenable. Our research offers compelling reasons to continue exploring the potential of research analytics to advance innovative methodologies and optimize the quality and impact of digital health care.

Acknowledgments

The authors are grateful to the iOUCH research staff for their commitment and contribution to this research. The authors also wish to thank Dr Chitra Lalloo for reviewing an early draft of this manuscript and providing comments and Mr Ryan Ho for sharing his insights on the service of research analytics and designing the APEEE service blueprint.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Analytics Platform to Evaluate Effective Engagement (APEEE) service blueprint.
[PNG File , 353 KB - jmir_v21i11e14849_app1.png]

Multimedia Appendix 2
Micro level semistructured interview script.
[PDF File (Adobe PDF File), 67 KB - jmir_v21i11e14849_app2.pdf]

References

1. Arigo D, Jake-Schoffman DE, Wolin K, Beckjord E, Hekler EB, Pagoto SL. The history and future of digital health in the field of behavioral medicine. J Behav Med 2019 Feb;42(1):67-83. [doi: 10.1007/s10865-018-9966-z] [Medline: 30825090]
2. Sharma A, Harrington RA, McClellan MB, Turakhia MP, Eapen ZJ, Steinhubl S, et al. Using digital health technology to better generate evidence and deliver evidence-based care. J Am Coll Cardiol 2018 Jun 12;71(23):2680-2690 [FREE Full text] [doi: 10.1016/j.jacc.2018.03.523] [Medline: 29880129]

3. Michie S, Yardley L, West R, Patrick K, Greaves F. Developing and evaluating digital interventions to promote behavior change in health and health care: recommendations resulting from an international workshop. J Med Internet Res 2017 Jun 29;19(6):e232 [FREE Full text] [doi: 10.2196/jmir.7126] [Medline: 28663162]
4. Patrick K, Hekler EB, Estrin D, Mohr DC, Riper H, Crane D, et al. The pace of technologic change: implications for digital health behavior intervention research. Am J Prev Med 2016 Nov;51(5):816-824. [doi: 10.1016/j.amepre.2016.05.001] [Medline: 27745681]
5. Klasnja, Smith S, Seewald NJ, Lee A, Hall K, Luers B, et al. Efficacy of contextually tailored suggestions for physical activity: a micro-randomized optimization trial of HeartSteps. Ann Behav Med 2019 May 3;53(6):573-582 [FREE Full text] [doi: 10.1093/abm/kay067] [Medline: 30192907]
6. Nahum-Shani I, Smith SN, Spring BJ, Collins LM, Witkiewitz K, Tewari A, et al. Just-in-Time Adaptive Interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. Ann Behav Med 2018 May 18;52(6):446-462 [FREE Full text] [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]
7. Crane D, Garnett C, Michie S, West R, Brown J. A smartphone app to reduce excessive alcohol consumption: identifying the effectiveness of intervention components in a factorial randomised control trial. Sci Rep 2018 Mar 12;8(1):4384 [FREE Full text] [doi: 10.1038/s41598-018-22420-8] [Medline: 29531280]
8. Ford I, Norrie J. Pragmatic trials. N Engl J Med 2016 Aug;375(5):454-463. [doi: 10.1056/nejmra1510059]
9. Collins LM, Kugler KC. Optimization Of Behavioral, Biobehavioral, And Biomedical Interventions: The Multiphase Optimization Strategy (MOST). Cham, Switzerland: Springer; 2018.
10. Bates DW, Saria S, Ohno-Machado L, Shah A, Escobar G. Big data in health care: using analytics to identify and manage high-risk and high-cost patients. Health Aff (Millwood) 2014 Jul;33(7):1123-1131. [doi: 10.1377/hlthaff.2014.0041] [Medline: 25006137]
11. Pham Q, Graham G, Carrion C, Morita PP, Seto E, Stinson JN, et al. A library of analytic indicators to evaluate effective engagement with consumer mHealth apps for chronic conditions: scoping review. JMIR Mhealth Uhealth 2019 Jan 18;7(1):e11941 [FREE Full text] [doi: 10.2196/11941] [Medline: 30664463]
12. van Gemert-Pijnen JE, Kelders SM, Bohlmeijer ET. Understanding the usage of content in a mental health intervention for depression: an analysis of log data. J Med Internet Res 2014 Jan 31;16(1):e27 [FREE Full text] [doi: 10.2196/jmir.2991] [Medline: 24486914]
13. Murray JM, French DP, Patterson CC, Kee F, Gough A, Tang J, et al. Predicting outcomes from engagement with specific components of an internet-based physical activity intervention with financial incentives: process analysis of a cluster randomized controlled trial. J Med Internet Res 2019 Apr 19;21(4):e11394 [FREE Full text] [doi: 10.2196/11394] [Medline: 31002304]
14. Sieverink F, Kelders S, Poel M, van Gemert-Pijnen L. Opening the black box of electronic health: collecting, analyzing, and interpreting log data. JMIR Res Protoc 2017 Aug 7;6(8):e156 [FREE Full text] [doi: 10.2196/resprot.6452] [Medline: 28784592]
15. Miller S, Ainsworth B, Yardley L, Milton A, Weal M, Smith P, et al. A framework for Analyzing and Measuring Usage and Engagement Data (AMUsED) in digital interventions: viewpoint. J Med Internet Res 2019 Feb 15;21(2):e10966 [FREE Full text] [doi: 10.2196/10966] [Medline: 30767905]
16. Pham Q, Graham G, Lalloo C, Morita PP, Seto E, Stinson JN, et al. An analytics platform to evaluate effective engagement with pediatric mobile health apps: design, development, and formative evaluation. JMIR Mhealth Uhealth 2018 Dec 21;6(12):e11447 [FREE Full text] [doi: 10.2196/11447] [Medline: 30578179]
17. Wyatt JC. How can clinicians, specialty societies and others evaluate and improve the quality of apps for patient use? BMC Med 2018 Dec 3;16(1):225 [FREE Full text] [doi: 10.1186/s12916-018-1211-7] [Medline: 30501638]
18. Kvedar JC, Fogel AL. mHealth advances clinical research, bit by bit. Nat Biotechnol 2017 Apr 11;35(4):337-339. [doi: 10.1038/nbt.3851] [Medline: 28398316]
19. Raymond N. Safeguards for human studies can't cope with big data. Nature 2019 Apr;568(7752):277. [doi: 10.1038/d41586-019-01164-z] [Medline: 30988466]
20. Gostin LO, Halabi SF, Wilson K. Health data and privacy in the digital era. J Am Med Assoc 2018 Jul 17;320(3):233-234. [doi: 10.1001/jama.2018.8374] [Medline: 29926092]
21. Yin RK, Campbell DT. Case Study Research And Applications: Design And Methods. Thousand Oaks, California: Sage Publications Inc; 2019.
22. SPARK. URL: https://projectspark.ca/ [accessed 2019-04-25]
23. MaRS Discovery District. SPARK Website Unleashes Digital Health Data URL: https://www.marsdd.com/news/spark-website-unleashes-digital-health-data/ [accessed 2019-04-25]
24. People – Howell Lab. iOUCH Lab: Improving Outcomes in Child Health through Technology URL: https://lab.research.sickkids.ca/iouch/ [accessed 2019-04-25]
25. Stinson JN, Lalloo C, Harris L, Isaac L, Campbell F, Brown S, et al. iCanCope with Pain™: User-centred design of a web- and mobile-based self-management program for youth with chronic pain based on identified health care needs. Pain Res Manag 2014;19(5):257-265 [FREE Full text] [doi: 10.1155/2014/935278] [Medline: 25000507]
26. eHealth Innovation. iCanCope URL: http://ehealthinnovation.org/project/icancope/ [accessed 2019-04-25]

27. Greenhalgh T, Robert G, Macfarlane F, Bate P, Kyriakidou O. Diffusion of innovations in service organizations: systematic review and recommendations. Milbank Q 2004;82(4):581-629 [FREE Full text] [doi: 10.1111/j.0887-378X.2004.00325.x] [Medline: 15595944]
28. Greenhalgh T, Stramer K, Bratan T, Byrne E, Mohammad Y, Russell J. Introduction of shared electronic records: multi-site case study using diffusion of innovation theory. Br Med J 2008 Oct 23;337:a1786 [FREE Full text] [doi: 10.1136/bmj.a1786] [Medline: 18948344]
29. Greenhalgh T, Shaw S, Wherton J, Vijayaraghavan S, Morris J, Bhattacharya S, et al. Real-world implementation of video outpatient consultations at macro, meso, and micro levels: mixed-method study. J Med Internet Res 2018 Apr 17;20(4):e150 [FREE Full text] [doi: 10.2196/jmir.9897] [Medline: 29625956]
30. Hsieh H, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
31. Yin RK. The case study crisis: some answers. Adm Sci Q 1981;26(1):58. [doi: 10.2307/2392599]
32. Sandelowski M. The use of quotes in qualitative research. Res Nurs Health 1994 Dec;17(6):479-482. [doi: 10.1002/nur.4770170611] [Medline: 7972926]
33. Conklin J, Farrell B, Ward N, McCarthy L, Irving H, Raman-Wilms L. Developmental evaluation as a strategy to enhance the uptake and use of deprescribing guidelines: protocol for a multiple case study. Implement Sci 2015 Jun 18;10:91 [FREE Full text] [doi: 10.1186/s13012-015-0279-0] [Medline: 26084798]
34. Ministry of Health and Long-Term Care. 2009. Patients First: Action Plan for Health Care URL: http://www.health.gov.on.ca/en/ms/ecfa/healthy_change/docs/rep_patientsfirst.pdf [accessed 2019-09-24]
35. Canada Health Infoway. Ontario's Patients First: Digital Health Strategy URL: https://www.infoway-inforoute.ca/en/component/edocman/3198-keynote-ontario-s-digital-health-strategy/view-document?Itemid=0 [accessed 2019-09-24]
36. desLibris. The Catalyst Towards An Ontario Health Innovation Strategy URL: https://www.deslibris.ca/ID/245701 [accessed 2019-09-24]
37. Government of Ontario. 2016. Value and Opportunities Created by Ontario's Digital Health Assets URL: https://www.ontario.ca/page/value-and-opportunities-created-ontarios-digital-health-assets [accessed 2019-09-24]
38. Life Sciences Ontario. Life Sciences Ontario Reacts to Ontario Health Innovation Council (OHIC) Report URL: https://lifesciencesontario.ca/news/life-sciences-ontario-reacts-ontario-health-innovation-council-ohic-report/ [accessed 2019-04-25]
39. MaRS Discovery District. MaRS Responds to the Ontario Health Innovation Council (OHIC) Landmark 'Catalyst' Report URL: https://www.marsdd.ca/media-centre/mars-responds-ontario-health-innovation-council-innovation-report/ [accessed 2019-04-25]
40. CPI Global Summit 2018. MEDEC Strongly Supports Ontario Budget Initiatives to Increase Innovation in Healthcare URL: https://cdn.ymaws.com/medtechcanada.org/resource/resmgr/New_Releases_2009-2015/23_April_2015_medec_ontario_.pdf [accessed 2019-09-26]
41. Central West LHIN. 10-point Digital Health Action Plan URL: https://tinyurl.com/y3zqqwvd [accessed 2019-09-27]
42. Data catalogue | Ontario. Improving Care: The Present and Future of Digital Health in Ontario URL: https://files.ontario.ca/3._ehealth_centre_of_excellence.pdf [accessed 2019-09-27]
43. Ontario Centres of Excellence. Health Technologies Fund URL: https://www.oce-ontario.org/programs/commercialization/health-technologies-fund [accessed 2019-04-25]
44. Ministry of Health and Long-Term Care. Call for Applications for a Provincial Centre of Excellence (COE) in Digital Health Benefits Evaluation. URL: https://www.trentu.ca/researchinnovation/sites/trentu.ca.researchinnovation/files/documents/Research%20Matter%20Nov2017.pdf [accessed 2019-09-28]
45. Women's College Hospital. Healthcare Revolutionized 2018-2022: Learn How We're Living Our Strategy—goal 4 URL: https://www.womenscollegehospital.ca/news-and-events/Connect-2018/Healthcare-Revolutionized-2018-2022-Learn-how-were-living-our-strategy-Goal-4 [accessed 2019-04-25]
46. Government of Canada. Draft Guidance Document - Software as a Medical Device (SaMD) URL: https://www.canada.ca/en/health-canada/services/drugs-health-products/public-involvement-consultations/medical-devices/software-medical-device-draft-guidance/document.html [accessed 2019-04-16]
47. Government of Canada. LawsJustice. Food and Drugs Act URL: https://laws-lois.justice.gc.ca/eng/acts/f-27/ [accessed 2019-04-25]
48. LawsJustice. Medical Devices Regulations URL: https://laws-lois.justice.gc.ca/eng/regulations/sor-98-282/ [accessed 2019-04-25]
49. Government of Canada. Notice of Intent: Strengthening the Post-market Surveillance and Risk Management of Medical Devices in Canada URL: https://tinyurl.com/y4cto4by [accessed 2019-04-25]
50. Government of Canada. Strengthening the Use of Real World Evidence and Regulations for Medical Devices URL: https://tinyurl.com/y57vfmcp [accessed 2019-04-25]
51. Palermo TM, Zempsky WT, Dampier CD, Lalloo C, Hundert AS, Murphy LK, et al. iCanCope with Sickle Cell Pain: design of a randomized controlled trial of a smartphone and web-based pain self-management program for youth with sickle cell disease. Contemp Clin Trials 2018 Nov;74:88-96. [doi: 10.1016/j.cct.2018.10.006] [Medline: 30316999]

52. Pham Q, Wiljer D, Cafazzo JA. Beyond the randomized controlled trial: a review of alternatives in mHealth clinical trial methods. JMIR Mhealth Uhealth 2016 Sep 9;4(3):e107 [FREE Full text] [doi: 10.2196/mhealth.5720] [Medline: 27613084]
53. Government of Canada. Funding Decisions Database URL: http://webapps.cihr-irsc.gc.ca/decisions/p/project_details.html?applId=398791&lang=en [accessed 2019-04-25]
54. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
55. Denis J, Hébert Y, Langley A, Lozeau D, Trottier L. Explaining diffusion patterns for complex health care innovations. Health Care Manage Rev 2002;27(3):60-73. [doi: 10.1097/00004010-200207000-00007] [Medline: 12146784]
56. Shaw J, Agarwal P, Desveaux L, Palma DC, Stamenova V, Jamieson T, et al. Beyond 'implementation': digital health innovation and service design. NPJ Digit Med 2018;1:48 [FREE Full text] [doi: 10.1038/s41746-018-0059-8] [Medline: 31304327]
57. Murad MH, Asi N, Alsawas M, Alahdab F. New evidence pyramid. Evid Based Med 2016 Aug;21(4):125-127 [FREE Full text] [doi: 10.1136/ebmed-2016-110401] [Medline: 27339128]
58. Greenhalgh T, Russell J. Why do evaluations of eHealth programs fail? An alternative set of guiding principles. PLoS Med 2010 Nov 2;7(11):e1000360 [FREE Full text] [doi: 10.1371/journal.pmed.1000360] [Medline: 21072245]
59. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. J Med Internet Res 2017 Nov 1;19(11):e367 [FREE Full text] [doi: 10.2196/jmir.8775] [Medline: 29092808]
60. Greenhalgh T, Papoutsi C. Studying complexity in health services research: desperately seeking an overdue paradigm shift. BMC Med 2018 Jun 20;16(1):95 [FREE Full text] [doi: 10.1186/s12916-018-1089-4] [Medline: 29921272]
61. Gelatt HB. Future sense: creating the future. Futurist 1993;27(5):- [FREE Full text]
62. ResearchNet. 2018. Team Grant : Transitions in Care URL: https://www.researchnet-recherchenet.ca/rnr16/vwOpprtntyDtls.do?prog=2956 [accessed 2019-05-05]
63. Graham J. University Affairs: Canadian university news, opinion, careers. 2011. The End of Medical Anthropology in Canada? URL: https://www.universityaffairs.ca/opinion/in-my-opinion/the-end-of-medical-anthorpology-in-canada/#comments [accessed 2019-05-05]
64. Canadian Institutes of Health Research. 2018. Ethical Considerations About Health Research URL: http://www.cihr-irsc.gc.ca/e/48801.html [accessed 2019-05-05]
65. Canadian Institutes of Health Research. Innovative Clinical Trials Initiative URL: http://www.cihr-irsc.gc.ca/e/49773.html [accessed 2019-05-07]
66. ResearchNet. Catalyst Grant : SPOR Innovative Clinical Trials URL: https://tinyurl.com/y465ffeq [accessed 2019-05-14]
67. Wolpert M, Jacob J, Napoleone E, Whale A, Calderon A, Edbrooke-Childs J. CORC Child Outcomes Research Consortium. Child- and Parent-reported Outcomes and Experience from Child and Young People's Mental Health Services 2011–2015 URL: https://tinyurl.com/y34262y3 [accessed 2019-09-24]
68. Wolpert M, Rutter H. Using flawed, uncertain, proximate and sparse (FUPS) data in the context of complexity: learning from the case of child mental health. BMC Med 2018 Jun 13;16(1):82 [FREE Full text] [doi: 10.1186/s12916-018-1079-6] [Medline: 29895295]
69. Ministry of Health and Long-Term Care. 2018. Chronic Disease Prevention Guideline, 2018 URL: http://health.gov.on.ca/en/pro/programs/publichealth/oph_standards/docs/protocols_guidelines/Chronic_Disease_Prevention_Guideline_2018.pdf [accessed 2019-09-28]
70. Mills A, Durepos G, Wiebe E. Case study as a methodological approach. In: Encyclopedia of Case Study Research. Thousand Oaks, California: Sage Publications; 2013.

Abbreviations

APEEE: Analytics Platform to Evaluate Effective Engagement
CIHR: Canadian Institutes of Health Research
COE: Center of Excellence
EPR: electronic patient records
FUPS: flawed, uncertain, proximate, and sparse
iCT: Innovative Clinical Trials
iOUCH: Improving Outcomes in Child Health through Technology
MOHLTC: Ministry of Health and Long-Term Care
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 04.06.19; peer-reviewed by F Agakov, T Irizarry; comments to author 30.07.19; revised version received 05.08.19; accepted 02.09.19; published 11.11.19.

Please cite as:
Pham Q, Shaw J, Morita PP, Seto E, Stinson JN, Cafazzo JA
The Service of Research Analytics to Optimize Digital Health Evidence Generation: Multilevel Case Study
J Med Internet Res 2019;21(11):e14849
URL: http://www.jmir.org/2019/11/e14849/
doi:10.2196/14849
PMID:31710296

Original Paper

A Typology of Patients Based on Decision-Making Styles: Cross-Sectional Survey Study

Mary Anne FitzPatrick[1*], PhD; Alexandra Claudia Hess[2*], PhD; Lynn Sudbury-Riley[3*], PhD; Peter Johannes Schulz[4*], PhD

[1]School of Management and Marketing, Waikato Management School, University of Waikato, Hamilton, New Zealand

[2]School of Communication, Journalism and Marketing, Massey University, Auckland, New Zealand

[3]Management School, University of Liverpool, Liverpool, United Kingdom

[4]Institute of Communication and Health, Faculty of Communication Science, University of Lugano, Lugano, Switzerland

[*]all authors contributed equally

Corresponding Author:
Mary Anne FitzPatrick, PhD
School of Management and Marketing
Waikato Management School
University of Waikato
Hillcrest Rd, Hillcrest,
Hamilton, 3240
New Zealand
Phone: 64 +64 7 838 4477
Email: maryfitz@waikato.ac.nz

Abstract

Background: Although previous research shows broad differences in the impact of online health information on patient-practitioner decision making, specific research is required to identify and conceptualize patient decision-making styles related to the use of online health information and to differentiate segments according to the influence of online information on patient decision making and interactions with health professionals.

Objective: This study aimed to investigate patients' decision making in relation to online health information and interactions with health care practitioners. We also aimed to present a typology of patients based on significant differences in their decision making.

Methods: We applied a large-scale cross-sectional research design using a survey. Data, generated using a questionnaire that was administered by companies specializing in providing online panels, were collected from random samples of baby boomers in the United Kingdom, the United States, and New Zealand. The total sample comprised 996 baby boomers born between 1946 and 1964, who had used the internet in the previous 6 months to search for and share health-related information. Data were analyzed using hierarchical cluster analysis and confirmatory factor analysis, as well as one-way analysis of variance, chi-square tests, and paired sample *t* tests.

Results: Analyses identified 3 key decision-making styles that served as the base for 4 unique and stable segments of patients with distinctive decision-making styles: the Collaborators (229/996, 23.0%), the Autonomous-Collaborators (385/996, 38.7%), the Assertive-Collaborators (111/996, 11.1%), and the Passives (271/996, 27.2%). Profiles were further developed for these segments according to key differences in the online health information behavior, demographics, and interactional behaviors of patients. The typology demonstrates that collaborative decision making is dominant among patients either in its pure form or in combination with autonomous or assertive decision making. In other words, most patients (725/996, 72.8%) show significant collaboration in their decision making with health care professionals. However, at times, patients in the combination Autonomous-Collaborative segment prefer to exercise individual autonomy in their decision making, and those in the combination Assertive-Collaborative segment prefer to be assertive with health professionals. Finally, this study shows that a substantial number of patients adopt a distinctly passive decision-making style (271/996, 27.2%).

Conclusions: The patient typology provides a framework for distinguishing practice-relevant and addressable segments with important implications for health care practitioners, including better-targeted communication programs for patients and more successful outcomes for health care services in the long term.

(J Med Internet Res 2019;21(11):e15332) doi:10.2196/15332

KEYWORDS

internet; online health information; patient decision making; patient-practitioner interaction; patient segments; patient typology; baby boomers; patient education

Introduction

Background

Within the patient-practitioner interaction, decision making is a critical interactional process that can be affected markedly by patient-sourced online health information. Recognized as an ethical imperative in terms of respect for patient preferences, values, and circumstances [1] and as a means to better manage scarce health care resources [2], decision making has crucial consequences ranging from patient health outcomes [3] to health care costs [4]. Previous research has found broad differences in the impact of online health information on decision making [5-7]. However, researchers have paid little attention to identifying and conceptualizing specific patient decision-making styles associated with online health information. In addition, no research has yet segmented or profiled patient groups based on their decision-making styles.

Patient Decision Making and Online Health Information

Decision making is a specific interactional process influenced by patient-sourced online health information [5,6,8-10]. Of particular relevance to this research are previous studies that detail a variety of decision-making behaviors by internet-informed patients; in particular, behaviors involving health care professionals. Previous research confirms that online health information can empower patients to be more active participants in decision-making with practitioners [7,11]. Such studies show that internet-informed patients can possess both knowledge and treatment preferences before they interact with practitioners, making them better equipped to take a fuller and more participatory role in decision making.

Shared decision making, the collaborative process in which available information is actively shared and health care decisions are made jointly by patients and practitioners, is "the focal point of a whole set of interlinking shifts and reforms related to the changing roles and relationships between doctors and patients" [12]. Grounded generally in broader principles of consumer rights, shared decision making in the context of patient-practitioner interactions reflects both a global upsurge in need for chronic care [13] and the shift from professional paternalism in health care services [14]. However, several scholars caution that shared decision making is an ideal [15], a viewpoint supported by evidence of a variety of patient decision-making experiences that do not conform to the ideal of participative collaboration. For example, patients can be reluctant to collaborate with practitioners either for fear of being labeled *difficult* and consequently less likely to receive quality care or because they perceive that active participation would threaten the relationship with their practitioner [15]. Previous research also finds patients can be uncomfortable asking questions of their practitioners because they prefer to conform to the passive role for patient or to defer to the practitioner as the authority [15,16].

In addition, patients can find it difficult to participate in collaborative decision making when their health literacy is low [17], when they lack the necessary evaluative skills to work through the decision-making process [18], or when they are facing a serious health condition [14]. Shared decision making is likely to be especially problematic for chronic patients who represent a sizeable—and growing—proportion of the population and are expected to take more responsibility themselves for managing their illnesses [17]. Patients with chronic conditions have more frequent consultations [19] and source complex health information from multiple sources, including changing networks of different health professionals [20].

Online health information also enables patients to self-diagnose and/or self-treat without visiting a practitioner. Prior research indicates that making internet-informed decisions independently of a practitioner can improve patients' self-efficacy and reduce unnecessary visits to practitioners—changes that are in line with trends toward increased patient responsibility and self-management of health [17,19]. In contrast, other scholars draw attention to the increased health risks for patients who decide to manage their health care independently [6,21].

Furthermore, patients can introduce online health information that is opposite to or actively challenges practitioners' diagnoses, prognoses, treatment, and advice. This set of patients' behaviors encompasses online health information that patients misunderstand but insist on, as well as information that patients interpret correctly but that conflicts with practitioners' opinions or threatens their professional authority [22]. These behaviors can result in inappropriate use of health care resources when the practitioner accedes to requests for unsuitable tests or treatments. Practitioner resistance, on the contrary, can result in patient dissatisfaction linked to nonadherence [23], seeking a second opinion (offline or online), switching health care providers, or changing the treatment plan [7].

Other studies confirm some patients decide not to share the online health information they have acquired with practitioners [6,24,25]. Having informed themselves before the interaction, nonsharing patients can feel more empowered in their decision making because they are better able both to understand what the practitioner says [22] and to trust the practitioner when the online information verifies the practitioner's explanation [26].

Styles of Decision Making

In summary, previous research has found important differences in patient behavior involving online health information and decision making with practitioners. On the basis of the literature, this study proposed the following 3 distinguishable

decision-making styles to be examined further in exploratory research:

- Collaborative decision making: following this style, the patient uses online health information to collaborate with the health professional and work as an active partner in making decisions about treatment and management.
- Autonomous decision making: this style applies to the patient who uses online health information to make autonomous decisions without involving the health care practitioner.
- Assertive decision making: the patient using this style draws on online information to assert his or her own preferences for treatment and management and/or to oppose practitioner advice in decision making.

Rationale

This research focused on investigating the decision-making behavior of baby boomer patients born between 1946 and 1964, an age cohort of substantial size. Baby boomers will place greater demands on health care professionals and exert considerable pressure on health care systems as longevity increases, rates of age-related and chronic conditions worsen, and lifetime health care costs escalate [27]. An increasing number of baby boomers uses the internet to search for and share health information [28] and assume a more active role in their own health care compared with preceding generations [29]. These changing patterns in baby boomers' health care consumption highlight an urgent need to better understand their patient behavior—in particular, their decision making—as heavy users of health care services. Such knowledge will provide a sound base for implementing health care programs that better meet the needs of different groups of internet-informed patients.

Objectives

The overall aim of this research was to address gaps in knowledge on patient decision making. We set out with the following 2 specific objectives: (1) to investigate the decision-making styles of patients in relation to online health information and interactions with health care practitioners and (2) to develop a typology of patients based on similarities and differences in those decision-making styles.

From these objectives, we then formulated the following research questions: (1) What are patient decision-making styles in relation to online health information and their health care practitioners? and (2) How can internet-informed patients be segmented and described based on their decision-making styles?

Methods

Research Design

To answer the research questions and achieve the study objectives, we applied a large-scale cross-sectional research design using a survey to generate quantitative data. The survey was conducted in the United Kingdom, the United States, and New Zealand to enable cross-national segmentation of patients according to their decision making, online health information, and interactional behaviors. We followed the research practices of other researchers conducting cross-national segmentation, who contend that because of globalization, segmentation should take account of commonalities across countries as well as differences within countries [30-34]. The rapid pace of globalization and the spread of a global culture fueled by the internet [30,34] endorse cross-national segmentation as an appropriate approach to identifying between-country segments of internet-informed patients with similar decision-making styles. We measured the patient behaviors at the individual level (following Steenkamp and Ter Hofstede [35]) rather than at the country level and worked with a single dataset, as is standard in cross-national segmentation, to develop a segmentation scheme that would partition respondents from the 3 countries into groups of patients with similar decision-making styles. We used a *country of residence* variable to account for differences by country, enabling fuller descriptions and better characterization of patient segments. Ethics committees of the University of Waikato, the University of Liverpool, and the University of Lugano approved the research before it commenced.

Questionnaire Development

The first step involved developing items for the proposed patient decision-making styles associated with online health information seeking (collaborative, autonomous, and assertive) and patient-practitioner interactions. We adopted well-established procedures [36,37] to develop a multidimensional measure of patient decision-making styles. For example, following Boateng et al [37], we based item generation on a review of the literature and existing scales, combined with data from focus groups (3 groups of 6 participants each: 1 mixed gender group, 1 female, and 1 male) and semistructured in-depth interviews (8 participants).

A total of 22 items was generated relating to patient-sourced online health information and decision making involving the health care practitioner. To ensure content validity, other faculty members experienced in the subject area (ie, expert judges) rated how well each of the items related to the 3 proposed decision-making styles. The resulting instrument was then subjected to 3 rounds of pretesting for item reduction and scale purification purposes. Together, these processes resulted in the retention of 10 items: 3 collaborative items, 3 autonomous items, and 4 assertive items. All 10 items appear in Table 1 (8 items appear within the table, and 2 items are listed in the table footnote). The final 10 items relating to patient decision-making styles used a Likert agreement scale ranging from 1 to 5, in line with researchers who employ this scale as an ordinal approximation of a continuous variable without compromising data analysis [38-41].

Table 1. Items and factor loadings of exploratory factor analysis.

Patient behavior item[a]	Patient decision-making style		
	Collaborative	Autonomous	Assertive
I have used online health information to ask questions of my health professional(s)	0.75	—[b]	—
I have had my diagnosis confirmed by my health professional	0.78	—	—
I have sought help from a health professional	0.86	—	—
I have changed the treatment recommended by a health professional	—	—	0.82
I have refused or discontinued treatment recommended by my health professional	—	—	0.83
I have changed from one health professional to another	—	—	0.75
I have tried to treat a health condition or disease without help from a health professional	—	0.83	—
I have tried to diagnose a health condition or disease I or someone else might have	—	0.84	—
Eigenvalue	3.591	1.685	3.395

[a]On the basis of split loadings and reliability analysis, the following 2 items were removed: "I have sought a second opinion from another health professional" (assertive) and "I didn't need to visit a health professional" (autonomous).

[b]Loadings of less than 0.5 are not shown to improve readability.

Following common practice in cluster analysis, our questionnaire design went beyond investigation of the 3 decision-making styles (which we refer to as *active* variables) to include other variables (which we refer to as *passive* variables); thus, we could identify practice-relevant and addressable segments. For example, we included frequency of visits to a health professional (measured using items based on questions in the Health Information National Trends Survey 2014; refer [42]) as well as variables relating to patients' online behavior and online information outcomes (see Multimedia Appendix 1 for more details). In addition, we used a range of other demographic variables, including *country of residence* to allow for differences by country in adopting different decision-making styles (see Multimedia Appendix 2 for more details). We used health-related variables such as living with chronic health problems and electronic health (eHealth) literacy (measured using items based on questions in Pew Research Center's *The Internet and Health* questionnaire [43]). The questionnaire provided a *not applicable* response for suitable questions to ensure there would be no missing data.

To check validity and reliability of the final items, we again asked the 6 experts to evaluate how well the constructs we wanted to measure were represented by relevant items in the questionnaire. Then, we further confirmed the validity and reliability of the survey via a pilot test of the questionnaire for readability and understandability (n=64). The questionnaire was prepared and administered in English to avoid problems translating questions and issues with conceptual equivalence [30].

Procedure

We outsourced the recruitment of baby boomer patients to external commercial organizations that specialize in online panels [44]. The organizations were instructed to sample respondents born between 1946 and 1964 who had used the internet in the previous 6 months to search for and share health-related information. The organizations were directed to distribute the link to the online questionnaire to a representative sample in terms of gender, ethnicity, education, income, and location. For sampling, the companies relied on registered members in their databases who receive compensation (points based) for participating in individual surveys. Data were collected over a 4-month period in the United Kingdom, the United States, and New Zealand to facilitate cross-national segmentation. A total sample of 996 (the United Kingdom, n=407; the United States, n=313; and New Zealand, n=276) respondents completed usable questionnaires with no missing data.

Data Analysis

To enable us to explore the existence of a common segmentation scheme across the United Kingdom, the United States, and New Zealand, data from the 3 countries were aggregated for analysis. First, we conducted a set of preliminary analyses, including descriptive data analysis, identification of outliers, and nonnormality checks. Then, to empirically test our 3 decision-making styles, we conducted an exploratory factor analysis, followed by a confirmatory factor analysis (CFA). Following this, based on the 3 decision-making styles (active variables), we conducted a cluster analysis. We began by employing a hierarchical cluster procedure using Ward method and measured the distance between cases using the square of the Euclidean distance. This procedure was followed by a cluster analysis using k-means. To develop fuller profiles of each segment (following Barnes et al [45]), we used descriptive characteristics including online information outcome, health-related domain, online domain, and demographic variables as passive variables (see Multimedia Appendices 2-5 for details). We used one-way analysis of variance, chi-square tests, and paired sample *t* tests to describe our segments. Data analysis resulted in a typology comprising 4 distinct patient segments with multidimensional profiles, which we present in the following sections.

Results

Sample Description

The sample contained slightly more females than males (499/996, 50.1%). In terms of ethnicity, the majority of respondents identified themselves as white or Caucasian (872/996, 87.6% of total sample; 389/407, 95.6% of UK respondents; 249/313, 79.6% of US respondents; and 234/276, 84.8% of New Zealand respondents). In terms of educational levels, most respondents had completed college and practical, technical, or occupational training (350/996, 35.1%), followed by 33% (329/996) with a university degree, and 30.6% (305/996) reporting high school as the highest educational level (see Table 2 for details). The total sample appears to be relatively homogenous on these demographic details, and the results should be interpreted in light of this.

Table 2. Sample characteristics by country (N=996).

Characteristics	United Kingdom (n=407)	United States (n=313)	New Zealand (n=276)	Total (N=996)
Sex, n (%)				
Male	193 (47.4)	163 (52.1)	141 (51.1)	497 (49.9)
Female	214 (52.6)	150 (47.9)	135 (48.9)	499 (50.1)
Age (years), n (%)				
46-49	40 (9.8)	55 (17.6)	72 (26.1)	167 (16.8)
50-54	124 (30.5)	72 (23.0)	64 (23.2)	260 (26.1)
55-59	95 (23.3)	92 (29.4)	69 (25.0)	256 (25.7)
60-64	148 (36.4)	94 (30)	25.7 (52.1)	313 (31.4)
Ethnicity, n (%)				
White	389 (95.6)	249 (79.6)	234 (84.8)	872 (87.55)
Non-white	18 (94.4)	64 (20.4)	42 (15.2)	124 (12.45)
Educational attainment, n (%)				
Less than high school	3 (0.7)	8 (2.6)	0 (0.0)	12 (1.2)
High school	158 (38.8)	59 (18.8)	89 (32.2)	305 (30.6)
College/practical/technical/occupational	148 (36.4)	101 (32.3)	101 (36.6)	350 (35.1)
University degree	98 (24.1)	145 (46.3)	86 (31.2)	329 (33.0)

Patient Decision-Making Styles

EFA supported the 3 decision-making styles indicated by the literature and exploratory research (see Table 1 for factor loadings). On the basis of split loadings and reliability analysis, 2 of the 10 items were eliminated. The total variance extracted by the 3 factors was 72%, the Bartlett test of sphericity was significant with $P<.001$, and the Kaiser-Meyer-Olkin measure was 0.78, indicating construct validity. The Cronbach *alpha* values of the remaining 8 items were of an acceptable standard for the study's item structure (collaborative *alpha*=.77, autonomous *alpha*=.69, and assertive *alpha*=.77; refer to the study by Loewenthal [46]).

Furthermore, a CFA of the 8 items (applying AMOS 24 testing for convergent and discriminant validity) indicated that the 3 measurement models provided a good model fit (see Table 3). Factor loadings of all items in the 3 models were higher than 0.6. Average variance extracted (AVE) was greater than 0.5, and compositional reliability (CR) was above 0.6. Thus, together, AVE and CR satisfied the requirement for convergent validity [32]. Finally, AVEs were higher than the squared correlations between constructs, thus supporting the discriminant validity of the model (see Table 3 for details) [47].

Table 3. Descriptive statistics and convergent and discriminant validity results.

Construct[a]	Mean based on total sample (SD)	Parameter estimates of confirmatory factor analysis (range)	Compositional reliability	Average variance extracted	Correlation/discriminant validation			Statistics						
					Collaborative	Autonomous	Assertive	χ^2 (df=15)	*P* value	Goodness-of-fit index[b]	Comparative fit index[c]	Normed fit index[d]	Tucker-Lewis index[e]	Root mean square error of approximation[f]
Collaborative	2.83 (0.94)	0.61-1.0	0.836	0.64	—[g]	0.1	0.16	120.8	<.001	0.97	0.96	0.95	0.92	0.08
Autonomous	3.34 (1.0)	0.68-0.77	0.69	0.528	0.31	—	0.43	—	—	—	—	—	—	—
Assertive	3.81 (0.84)	0.61-0.82	0.783	0.549	0.4	0.18	—	—	—	—	—	—	—	—

[a]The calculated values of the square correlation coefficient between all possible pairs of constructs are presented in the upper triangle of the matrix. Correlations between all pairs of constructs are presented in the lower triangles of the matrix ($P<.01$).

[b]Values higher than 0.95 indicate better model fit [48].

[c]Values greater than 0.95 indicate a good fit [49].

[d]Values greater than 0.95 indicate a good fit.

[e]Values greater than 0.9 are considered a satisfactory fit [50].

[f]Values between 0.05 to 0.1 are considered a fair fit [51], with values below 0.08 acceptable [50].

[g]Not applicable.

Patient Segments

Initial analysis of data generated by the survey focused on the validation and interpretation of segments distinguished by statistical differences in the decision making of segment members. To determine the number of segments, first, a hierarchical cluster procedure (Ward method) was used, and the square of the Euclidean distance measured the distance between cases. Moreover, 3-, 4-, 5-, and 6-cluster solutions were explored. To determine the number of groups, the researchers considered the dendrograms and the distance at which each cluster was formed, profiled each cluster, and applied practical judgment and theoretical foundations [52]. All these indicators suggested that the 4-cluster solution was the most acceptable. Next, a k-means clustering analysis of the 4-cluster solution was performed. A pairwise cluster validation revealed that all 3 patient decision-making styles were significant in distinguishing between the 4 clusters. A pairwise analysis of the differences between the clusters revealed that, on average, 83% of the individual pairwise differences were significant at $P<.05$, showing the 4 clusters obtained using the patient decision-making styles are unique and stable. On the basis of analysis, we refer to the clusters as *segments* of patients labeled the Collaborators, the Autonomous-Collaborators, the Assertive-Collaborators, and the Passives.

Thus, as expected, we did find the 3 decision-making styles we derived from previous research. Furthermore, based on our cluster analysis, we found that the majority of patients adopt a Collaborative decision-making style (as we had proposed), including 2 segments that adopt combination Autonomous-Collaborator and Assertive-Collaborator decision-making. We also found 1 group of patients that does not adopt collaborative decision making: the segment of Passive patients (see Table 4 for details of each segment).

Our data analysis to this point showed that patients could be segmented into 4 distinct segments using the 3 decision-making style variables as active variables (see Table 4). In the next step, we used passive variables in our analysis, which allowed us to take account of differences (eg, country of residence) to provide fuller, more nuanced profiles of the 4 decision-making segments. Multimedia Appendices 2 and 3 provide descriptions of the segments by relevant variables.

The following subsections detail each segment. Full profiles of each are presented in the Discussion section.

Table 4. Final cluster solution and analysis of results.

Decision-making style	Total sample	Segment 1[a]	Segment 2	Segment 3	Segment 4	*F* test	*P* value
		Collaborators	Autonomous-Collaborators	Assertive-Collaborators	Passives		
Collaborative, mean (SD)	3.17 (0.94)	3.7 (0.55)	3.5 (0.65)	3.82 (0.59)	1.99 (0.55)	494.107 (995)	<.001
Autonomous, mean (SD)	2.61 (1.00)	1.98 (0.65)	3.56 (0.56)	2.28 (0.62)	1.96 (0.79)	435.242 (995)	<.001
Assertive, mean (SD)	2.2 (0.84)	1.66 (0.46)	2.64 (0.75)	3.13 (0.57)	1.63 (0.52)	287.477 (995)	<.001
Cluster size	__[b]	229	385	111	271	—	—
Percentage of respondents	—	23.0	38.7	11.1	27.2	—	—

[a]All segment means are significant at the .001 level, and 83% of the pairwise comparison is significant at *P*<.05 level. All variables are coded on a 5-point scale, with 1 for strongly disagree to 5 for strongly agree.

[b]Not applicable.

Segment 1: Collaborators

This segment describes 23.0% (229/996) of respondents, making it the third largest of the 4 segments. This research finds patients in segment 1 report that they rely significantly more on their health professional than the internet for health-related information. However, Collaborative decision makers state that as a result of searching online for health information, they can communicate with their health professionals significantly better compared with Autonomous-Collaborators (*P*<.05; see Multimedia Appendix 3 for more details). Collaborators, who are moderately eHealth literate (slightly lower than patients in the autonomous-collaborator segment with *P*=.08, that is, marginally significant; see Multimedia Appendix 4 for details), talk significantly more to practitioners about problems with the online health information they have accessed than Autonomous-Collaborator patients (*P*<.05; see Multimedia Appendix 5 for more details). In addition, Collaborators typically use the internet less to interact with others online about health-related matters than the combination Autonomous-Collaborators and Assertive-Collaborators (all *P* values <.05; see Multimedia Appendix 5 for more details). For example, collaborators go online less frequently to read others' health experiences, to share their own personal health experiences, or to post a comment or review online about a health-related product, service, or person.

Likely to be living with a chronic health problem (*P*<.05; see Multimedia Appendix 4 for more details), people in this segment visit their health care practitioners more often compared with Autonomous-Collaborators (*P*<.05; see Multimedia Appendix 3 for more details). This segment is not dominated by any particular nationality, with US citizens representing 34.9% (80/229), UK citizens representing 31.9% (73/229), and NZ citizens representing 33.2% (76/229). Segment 1 patients are most likely to be retired and receiving government welfare (eg, pension). This segment, which has the largest number of widowed respondents (19/53), includes a cross-section of ages across the baby boomer cohort (all *P* values <.05; see Multimedia Appendix 2 for more details).

Segment 2: Autonomous-Collaborators

The largest single group at 38.7% (385/996), the Autonomous-Collaborator segment, is made up of patients who exhibit a combination of both collaborative and autonomous styles in their decision making with health care practitioners. Autonomous-Collaborator patients differ significantly from Collaborators in that interaction and support from others online are more important, read others' commentaries about health online, and post more questions and reviews online (all *P* values <.05; see Multimedia Appendix 5 for more details). Patients in this segment frequently go online in an attempt to diagnose a health condition themselves. In fact, autonomous-collaborators report that they rely significantly more on the internet than the health professional for health-related information (*P*<.05). When they encounter a problem with internet-sourced health information, Autonomous-Collaborators talk about it with friends and/or someone online significantly more than Collaborators (*P*<.05; see Multimedia Appendix 5 for more details). This research shows that patients who use a combination of autonomous and collaborative decision-making styles in interactions with health care practitioners are slightly more eHealth literate than the pure Collaborator (*P*=.08; ie, marginally significant; see Multimedia Appendix 4 for more details). They are neither more nor less likely to be living with a chronic health problem (see Multimedia Appendix 4 for more details), and they make a moderate number of visits to health care practitioners (significantly less compared with Collaborators and Assertive-Collaborators; *P*<.05; see Multimedia Appendix 3 for more details).

In terms of other outcome variables, this segment reports slightly improved communication and relationship quality with their practitioners, yet less overall compared with Collaborator and Assertive-Collaborator patients (*P*<.05 difference for communication; *P*=0.05 difference in relationship quality between Autonomous-Collaborator and Assertive-Collaborator; see Multimedia Appendix 3 for more details). This Autonomous-Collaborator segment is comprised mainly of UK citizens; with 49.1% (189/385) of respondents, this is the largest percentage of any nationality in the research, whereas New

Zealand (81/385, 21%) and US (115/385, 29.9%) respondents represent significantly smaller national groups in the segment. Segment 2 patients are most likely to have never married. This segment is made up predominantly of patients born between 1960 and 1964, the Fourth Wave and youngest of the baby boomer generation (P<.05; see Multimedia Appendix 5 for more details).

Segment 3: Assertive-Collaborators

This segment, the smallest at 11.1% (111/996) of the sample, comprises patients characterized by a distinctive combination of collaborative and assertive decision-making styles. Assertive-Collaborators have the highest frequency of online searches for health-related information overall.

Data analysis reveals that Assertive-Collaborator patients are differentiated from the Autonomous-Collaborator and Collaborator segments in their interactional behavior. For instance, Assertive-Collaborators use their health professionals as a source of health information significantly more often than Autonomous-Collaborators (P<.05) and find the health professional more useful than the internet compared with Collaborators (P<.05; see Multimedia Appendix 5 for more details).

Strong social connections are reflected in results showing Assertive-Collaborators rate significantly more highly than Collaborators on a range of interactions with others, including sharing personal experiences online and contacting someone online when they have problems with online health information (all P values<.05; see Multimedia Appendix 5 for more details). The frequency of their online information searches on behalf of friends and coworkers is also higher than Collaborators (all P values<.05). Finally, Assertive-Collaborators search online for health information for family members more frequently than Autonomous-Collaborators (P<.05).

Assertive-Collaborator patients are highly eHealth literate and are more likely to be chronically ill (P<.05; see Multimedia Appendix 4 for more details) and to have a relatively high number of visits to health care practitioners (compared with the Autonomous-Collaborator; P<.05; see Multimedia Appendix 3 for more details). Interestingly, the Assertive-Collaborator patient also feels that both communication effectiveness and the relationship with the health care professional have improved markedly, with the highest means on both of these outcome variables across all 4 segments (see Multimedia Appendix 3 for more details). Results show Assertive-Collaborators are the oldest of the baby boomer cohort (ie, First Wave boomers born between 1946 and 1949), widowed or permanently separated, and mostly US citizens (48/111, 43.2%; P<.05; see Multimedia Appendix 2 for more details).

Segment 4: Passives

This segment, the second largest group comprising 27.2% (271/996) of respondents, is made up of patients who typically are not collaborative, assertive, or autonomous in decision making involving health care practitioners. Of all 4 segments, patients who are passive decision makers rate online health information the least useful in comparison with information provided by a health practitioner. More explicitly, compared with the other 3 patient segments, Passives report the internet is significantly less useful as a source of information in their health-related decision making (P<.05; see Multimedia Appendix 5 for more details). Passive patients (compared with the other segments) are less likely to have looked online to try to diagnose a health condition themselves, to have researched a health-related product or service online, and/or to have signed up to receive email updates or alerts online (all P values<.05; see Multimedia Appendix 4 for more details).

Our analysis finds these Passive patients are less likely to have chronic health problems (P<.05), they have low eHealth literacy compared with the other segments (P<.05; see Multimedia Appendix 4 for more details), and they do not visit their health care practitioners often compared with the other segments (P<.05; see Multimedia Appendix 3 for more details). They also report, as a result of accessing health information online, the least effect on the quality of communication and the lowest levels of relationship quality with their health care professionals (P<.05; see Multimedia Appendix 3 for more details). In terms of demographic variables, segment 4 patients are most likely Third Wave boomers born between 1955 and 1959, are in a dissolved union, or never married. UK patients dominated this segment (116/271, 42.8%; P<.05; see Multimedia Appendix 2 for more details).

Discussion

Principal Findings

The overarching purpose of this study was to investigate more closely the decision-making styles of patients who seek online health information and to develop a typology of patients based on significant differences in those styles. Initially, we proposed 3 decision-making styles in relation to health professionals: collaborative, autonomous, and assertive. Data analysis confirmed these 3 patient decision-making styles. However, analysis revealed that patients' decision making is considerably more complex than the 3 singular styles we proposed, with 2 combination styles of decision making and a fourth style distinguished by patient passivity. Specifically, to answer our first research question (see section Objectives), this research found patients can be clustered in 4 distinct segments; namely, Collaborators, combination Autonomous-Collaborators and Assertive-Collaborators, and Passive patients. In relation to the second research question, the research developed a typology of patients and described the segments based on similarities and differences in patient decision-making styles.

In summary, the major contributions of this paper include our identification of fundamental and distinct patient decision-making styles involving online health information, the typology of patients the empirical research revealed, and the multidimensional profiles (following) of those 4 patient segments developed on the basis of their decision making and other key practice-relevant variables.

Profile of Collaborators

Patients in the Collaborator segment are most likely to use the collaborative decision-making style and, according to the literature, participate with their practitioners in joint decision

making. According to our data analysis, patients who are collaborative decision makers actively seek help from health care professionals, they use online health information to ask questions in interactions with their practitioners, and they have diagnoses they have made themselves confirmed by a professional (see Table 1). Applying previous findings (see the study by Essén et al [53]), Collaborators can be expected to be active partners in managing their health care, taking responsibility for finding and appraising relevant online health information, for disclosing their perspectives and preferences, and for weighing treatment options. As a result of sharing with their health care professionals the online health information they have accessed, these patients feel they can ask better questions and participate in discussion at a deeper level. Consistent with collaborative decision making, these patients review information and options together with their practitioners and thus are more involved in decision making [7,9,10,16,54]. In addition, according to our data, Collaborators are influenced by online health information to seek professional help, to have diagnoses they have made confirmed by their health professional, and to find out more from the practitioner about health conditions, treatment, and/or management.

Although a lower than average number of these patients live with chronic health problems, collaborative decision makers visit their health professionals more often. Some researchers explain this distinctive pattern as being motivated by the patients wanting to treat conditions promptly and avoid further complications [55]. However, others warn that such behavior can be problematic when it results in overuse of the health care system [56]. Patients in the Collaborator segment are moderately eHealth literate compared with Autonomous-Collaborators and Assertive-Collaborators (both segments highly eHealth literate) and passives (low eHealth literacy). Building on the work by Schulz et al [28], more research could further investigate the relationships between chronic health problems, visits to health care professionals, and eHealth literacy. For example, 1 important clinical implication is that raising the eHealth literacy of patients in the Collaborators segment could enable more self-responsibility thereby reducing overdependence on health professionals and unnecessary visits to those practitioners.

Profile of Autonomous-Collaborators

The Autonomous-Collaborator segment includes patients who have accessed online health information and use a combination of decision-making styles: they are both autonomous and collaborative in their decision making with health professionals. Thus, at times, they make decisions in collaboration with practitioners (see section Segment 1 profile above), and at other times, they make decisions independently of health care practitioners; for instance, they try to diagnose a health condition or disease they or someone they know has, and they will also treat a health condition or disease without help from a health professional (Table 2).

When patients in this segment use autonomous decision making, they are self-active in addressing their health needs [7,16]. For example, autonomous decision makers access health information online to understand a specific condition, to explore whether symptoms are related to clinically meaningful diseases, and to diagnose and treat health conditions themselves [5,16]. In other words, patients practicing the Autonomous-Collaborative decision-making style will make autonomous decisions regarding their health at times without consulting practitioners.

Internet-informed patients can possess both knowledge and treatment preferences before they interact with practitioners; therefore, they are better equipped to take a fuller and more participatory role in decision making. Thus, in their interactions with health care professionals, these patients can choose to share information to be interpreted, evaluated, contextualized, and deliberated as part of collaborative decision making with practitioners. However, at other times, Autonomous-Collaborative decision makers will decide not to share their online health information; yet it is likely that having informed themselves before the interaction they feel more empowered in their decision making because they are better able to understand what the practitioner says or to trust the practitioner when the online information verifies the practitioner's explanation.

This single segment contains all respondents who practice autonomous decision making; however, the results clearly show that at times, these patients prefer to make joint decisions with practitioners. Similar to patients in the Assertive-Collaborator segment, patients who combine autonomous and collaborative decision making are highly eHealth literate. However, there are important differences between these 2 segments in terms of how the online health information they access influences patient decision making with health professionals and the frequency of visits to practitioners. Assertive-Collaborators typically challenge their practitioners based on the online information, whereas Autonomous-Collaborators incorporate the online information they have accessed independently in interactions with practitioners. In other words, although results indicate this segment uses online information to make decisions independently of health professionals (autonomous decision-making style), data show they collaborate and engage when they do consult practitioners. Interestingly, Assertive-Collaborators are more likely to have chronic health problems, which might explain the high number of visits to their health professionals. A question that arises here is whether the assertive decision-making style is impacted by chronic health problems and the potential frustration associated with being persistently sick. Future research could further investigate those relationships.

At a time when shared decision making and self-responsible/self-managing patients are seen as key to reducing the costs of health care service provision, Autonomous-Collaborators are clearly a desirable segment. Certainly, their collaborative behaviors are likely to make this segment efficient for health professionals to serve. These patients are likely to be easy and fast to work with; therefore, relational labor costs for the practitioner will be minimal. Therefore, from the perspective of health care policy makers, this segment (predominantly UK Fourth Wave baby boomers born between 1960 and 1964) will impose the smallest financial cost on the health care system. Future research on the implications of these findings for clinicians could identify and explain the reasons some patients are assertive and others are

autonomous in relation to their practitioners when making decisions.

Profile of Assertive-Collaborators

The Assertive-Collaborator segment comprises patients who will be both assertive and collaborative in decision making with practitioners. Thus, some of their health care decisions are collaborative in that these patients participate actively with their health care professionals in joint decisions on treatment and health management. However, these same patients are also self-determined decision makers who change, refuse, or discontinue recommended treatment and switch health professionals (Table 2).

All patients who are likely to be assertive and/or challenging with online health information belong to this specific segment. When the online health information they have sourced differs from that provided by practitioners, Assertive-Collaborators can use that information to challenge or oppose the opinions or recommendations of their practitioners and assert their own preferences for health care treatment. In line with previous research, patients in this segment can be expected to have contrary views to their practitioners, contradict practitioner interpretations of their health situation, and/or insist on tests and treatments based on online health information they have gathered [16]. When patients are misinformed by online health information, these tests and treatments are likely to be inappropriate. Conversely, when patients are better informed than practitioners, their health-related requests may well be appropriate. Finally, when a patient chooses to raise accurate online health information that conflicts with their practitioner's opinions, that decision can threaten the professional authority of the health care professional [22]. In summary, assertive patients can be challenging—sometimes confrontational—in their decision making with practitioners. In line with the literature [7], this research used the following as indicators of assertive patient responses: changing the treatment recommended by a health care practitioner, refusing or discontinuing recommended treatment, and changing from one professional to another.

At other times, these same patients are also collaborative in their decision making with practitioners. Therefore, an important clinical implication of these results is while their health professionals can be confident that this segment will challenge them at times, practitioners can also expect, in other circumstances, this segment will engage in joint decision making in which health information sourced online is shared and deliberated during the interaction. Health professionals can also be encouraged by the fact that these patients are displaying assertive decision-making behaviors inside the relationship rather than independently of the practitioner. This indicates that the nature of the relationship allows for debate, and that while online health information is important to the patient in their decision-making, so too is the practitioner.

Clinical implications for the Assertive-Collaborator segment include ensuring these patients have access to high-quality health information and that they maintain high eHealth literacy (the mean eHealth literacy of this segment is 2.19, highest shared with Autonomous-Collaborators). These strategies could reduce the potential for this segment to be misinformed, leading to confrontational interactions and decision making in the context of the patient-practitioner relationship, thus avoiding the possibilities of adverse health effects for the patient and of patient switching for the practitioner. Health professionals also could be trained in managing interactions with these patients so the online health information they share with practitioners is discussed and considered in a respectful open dynamic. In this way, practitioners' acknowledgment of their patients' rights and preferences in seeking and sharing online information would support such information activity as an integral part of patients taking more responsibility for their own health. For example, as a basis for shared decision making, early in each encounter, the practitioner could encourage the patient to talk about any health information they have accessed online preconsultation.

In addition, health professionals could be trained in conflict management and resolution so that the characteristic debate of Assertive-Collaborators' decision making can be vigorous without compromising the safety of the patient-practitioner relationship in which it occurs. This is the only segment that contains potentially confrontational patients; future studies should investigate the health-related situations in which this segment is most likely to be assertive and those when patients will be collaborative with health professionals. Finally, most of this segment is likely to suffer from a chronic health condition; more research is warranted on the links between specific health problems, online health information, and the decision making of Assertive-Collaborators.

Profile of Passives

Patients in the Passive segment characteristically are not influenced significantly by online health information to be collaborative, autonomous, or assertive in their decision making with health professionals. Instead, this group is characteristically passive, a finding consistent with previous work finding some patients prefer to rely on the traditional paternalistic patient-physician relationship and leave decision making to their practitioners [14,18]. Patients in this segment visit their health professionals the least frequently of all 4 segments in the typology. This segment (predominantly First Wave baby boomers from the United Kingdom) could be regarded as a risk group because coupled with the negligible impact of internet-accessed health information, the low number of health professional visits means there is the risk that serious conditions are undetected. Given the low influence of online health information on the decision making of this sizeable segment, policy makers and practitioners urgently need to identify those communication channels that are most effective in reaching and supporting these patients. Future research should establish the interaction and communication preferences of this segment to further develop the clinical implications for this segment.

Contributions

To conclude this section, it is worth noting that patients in both the Collaborator segment and the Assertive-Collaborator segment report chronic health problems. In addition, patients in both segments make more visits to their health care practitioners. Yet, the Assertive-Collaborator is significantly more eHealth literate than the Collaborator, suggesting a link

between eHealth literacy and more assertive patient behaviors in interactions with health care practitioners (as behavioral dimensions of patient empowerment). These results are in line with the recent findings of Schulz et al [28] that the number of visits to a health care professional is independent of eHealth literacy, and there is a significant positive relationship between high eHealth literacy and empowerment. For relatively healthy people, high eHealth literacy is linked with being both autonomous and collaborative in decision making with practitioners and with a medium number of practitioner visits (see section Profile of Autonomous-Collaborators). In contrast, healthy patients with lower levels of eHealth literacy are unaffected by online health information and have low frequency of visits to health care practitioners (see section Profile of Passives).

In addition, it is important to note that the majority of patients who use online health information in their decision making with health care professionals are characteristically collaborative to some degree, as reflected in the segment titles (Collaborators, Autonomous-Collaborators, and Assertive-Collaborators). For the 72.8% (725/996) of respondents in these 3 segments, collaboration with their practitioners is an important interactional behavior. In other words, the majority of patients show significant collaboration in their decision making with health care professionals. However, at times, 1 distinguishable group of that majority prefers to exercise individual autonomy in their decision making, and another group prefers to be assertive with their health professionals. In addition, this study finds that a substantial number of patients adopt a distinctly passive decision-making style (271/996, 27.2%) with practitioners.

Thus, the research makes 3 main contributions to the literature. First, it increases our understanding of online health information and patient decision making by identifying in the literature and exploratory research 3 patient decision-making styles: collaborative, autonomous, and assertive. Second, the research extends the literature on patient segments [57] by deriving a typology based on decision-making styles, additional sociodemographic and behavioral characteristics, and outcome variables. The 4 segments are profiled as the Collaborators, the Autonomous-Collaborators, the Assertive-Collaborators, and the Passive patients. Third, the research demonstrates that collaborative decision-making is dominant among patients, either in its pure form (229/996, 23.0%) or in combination with autonomous (385/996, 38.7%) or assertive decision making (111/996, 11.1%).

Limitations

This study, as any other study, has some limitations that offer opportunities for future research. First, the purpose of conducting this research in multiple countries was to identify shared patterns in patient decision making across the United Kingdom, the United States, and New Zealand. However, to allow for differences in data from the 3 countries, we included *country of residence* as one of the passive variables in our analysis. As discussed, we found significant associations between the country of residence of our respondents and the segments they belong to. We believe this considerably strengthens the contribution of our research. We also acknowledge that in this research, we cannot make any conclusions beyond the 3 western countries we investigated. This is important to note as patients in nonwestern countries with different core cultural values might have differences in their use of internet-sourced health information and patient decision making. More research is needed to replicate or identify differences across different cultures. Next, this research uncovered general patient decision-making styles. However, there might be numerous contextual factors that influence actual decision making in a specific context. Future research could investigate those factors and their impact on behavior (or a good proxy for that). In our research we deliberately focused on online information and did not account for other sources of health information (ie, books). It would be interesting to further investigate whether the same decision-making styles and typology emerge if patients use more traditional health information sources. In addition, this study looked at the behavior and decision making from the patient's perspective. One very interesting avenue for future research is to investigate how practitioners perceive different patient segments and how practitioner behaviors impact such segments. In addition, our sample was baby boomers, certainly an important cohort because of the effects on society of managing their health care; however, the typology presented here should be investigated using patients from other age groups. Furthermore, the research sample was fairly homogenous. Therefore, the results and their generalizability should be considered in light of the homogenous sample we have. Moreover, because we used cross-sectional data to study patients' decision-making styles, our research design did not account for the dynamic of variables. Moreover, we relied on self-reported outcomes that were not objectively measured; that is, they were not a measurement of how patients actually engage physicians during consultations. Finally, the way the survey was conducted might have influenced the results. A longitudinal study with real live data would be very useful as a follow-up. We believe that this research opens up a wide range of possibilities for future research into decision making and patient-practitioner interactions.

Conclusions

This research demonstrates complex differences regarding the decision making, online health information, and interactional behaviors of baby boomer patients. Close investigation of the variations in characteristics showed that according to their decision-making style, patients experience significant differences in the way online health information impacts their interactions with health professionals. The typology of patients based on their decision making provides a more sophisticated framework than simplistic descriptors (eg, *empowered* or *nonempowered* patients) or demographic characteristics (eg, age or nationality) for distinguishing practice-relevant and addressable segments of patients. Understanding these segments and their clinical implications will enable practitioners and policy makers to implement health care communication programs that are meaningful and valued by different groups of patients, thereby supporting more accurate targeting and more successful outcomes for health care services in the long term.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Questions/items for online domain and online information outcome variables.
[DOCX File , 19 KB - jmir_v21i11e15332_app1.docx]

Multimedia Appendix 2
Segments described by demographic variables.
[DOCX File , 20 KB - jmir_v21i11e15332_app2.docx]

Multimedia Appendix 3
Segments described by online information outcome variables.
[DOCX File , 19 KB - jmir_v21i11e15332_app3.docx]

Multimedia Appendix 4
Segments described by health-related domain variables.
[DOCX File , 19 KB - jmir_v21i11e15332_app4.docx]

Multimedia Appendix 5
Segments described by online domain variables.
[DOCX File , 27 KB - jmir_v21i11e15332_app5.docx]

References

1. Hoffmann TC, Del Mar CB. Shared decision making: what do clinicians need to know and why should they bother? Med J Aust 2014 Nov 3;201(9):513-514. [doi: 10.5694/mja14.01124] [Medline: 25358570]
2. Townsend A, Leese J, Adam P, McDonald M, Li LC, Kerr S, et al. eHealth, participatory medicine, and ethical care: a focus group study of patients' and health care providers' use of health-related internet information. J Med Internet Res 2015 Jun 22;17(6):e155 [FREE Full text] [doi: 10.2196/jmir.3792] [Medline: 26099267]
3. Stiggelbout AM, van der Weijden T, de Wit MP, Frosch D, Légaré F, Montori VM, et al. Shared decision making: really putting patients at the centre of healthcare. Br Med J 2012 Jan 27;344:e256. [doi: 10.1136/bmj.e256] [Medline: 22286508]
4. Elwyn G, Frosch DL, Kobrin S. Implementing shared decision-making: consider all the consequences. Implement Sci 2016 Aug 8;11:114 [FREE Full text] [doi: 10.1186/s13012-016-0480-9] [Medline: 27502770]
5. Anderson JG, Rainey MR, Eysenbach G. The impact of CyberHealthcare on the physician-patient relationship. J Med Syst 2003 Feb;27(1):67-84. [doi: 10.1023/a:1021061229743] [Medline: 12617199]
6. Imes RS, Bylund CL, Sabee CM, Routsong TR, Sanford AA. Patients' reasons for refraining from discussing internet health information with their healthcare providers. Health Commun 2008 Nov;23(6):538-547. [doi: 10.1080/10410230802460580] [Medline: 19089701]
7. Tan SS, Goonawardene N. Internet health information seeking and the patient-physician relationship: a systematic review. J Med Internet Res 2017 Jan 19;19(1):e9 [FREE Full text] [doi: 10.2196/jmir.5729] [Medline: 28104579]
8. Fogel J, Albert SM, Schnabel F, Ditkoff BA, Neugut AI. Use of the Internet by women with breast cancer. J Med Internet Res 2002;4(2):E9 [FREE Full text] [doi: 10.2196/jmir.4.2.e9] [Medline: 12554556]
9. Gerber BS, Eiser AR. The patient physician relationship in the internet age: future prospects and the research agenda. J Med Internet Res 2001;3(2):E15 [FREE Full text] [doi: 10.2196/jmir.3.2.e15] [Medline: 11720957]
10. Murray E, Lo B, Pollack L, Donelan K, Catania J, White M, et al. The impact of health information on the internet on the physician-patient relationship: patient perceptions. Arch Intern Med 2003 Jul 28;163(14):1727-1734. [doi: 10.1001/archinte.163.14.1727] [Medline: 12885689]
11. Shay LA, Lafata JE. Where is the evidence? A systematic review of shared decision making and patient outcomes. Med Decis Making 2015 Jan;35(1):114-131 [FREE Full text] [doi: 10.1177/0272989X14551638] [Medline: 25351843]
12. Cribb A, Donetto S. Patient involvement and shared decision-making: an analysis of components, models and practical knowledge. Eur J Pers Cent Health 2013;1(1):41-49. [doi: 10.5750/ejpch.v1i1.634]
13. Tinetti ME, Fried TR, Boyd CM. Designing health care for the most common chronic condition--multimorbidity. J Am Med Assoc 2012 Jun 20;307(23):2493-2494 [FREE Full text] [doi: 10.1001/jama.2012.5265] [Medline: 22797447]
14. Charles C, Gafni A, Whelan T. Decision-making in the physician-patient encounter: revisiting the shared treatment decision-making model. Soc Sci Med 1999 Sep;49(5):651-661. [doi: 10.1016/s0277-9536(99)00145-8] [Medline: 10452420]

15. Frosch DL, May SG, Rendle KA, Tietbohl C, Elwyn G. Authoritarian physicians and patients' fear of being labeled 'difficult' among key obstacles to shared decision making. Health Aff (Millwood) 2012 May;31(5):1030-1038. [doi: 10.1377/hlthaff.2011.0576] [Medline: 22566443]
16. Sommerhalder K, Abraham A, Zufferey MC, Barth J, Abel T. Internet information and medical consultations: experiences from patients' and physicians' perspectives. Patient Educ Couns 2009 Nov;77(2):266-271. [doi: 10.1016/j.pec.2009.03.028] [Medline: 19411157]
17. Heijmans M, Waverijn G, Rademakers J, van der Vaart R, Rijken M. Functional, communicative and critical health literacy of chronic disease patients and their importance for self-management. Patient Educ Couns 2015 Jan;98(1):41-48. [doi: 10.1016/j.pec.2014.10.006] [Medline: 25455794]
18. Charles C, Gafni A, Whelan T. Shared decision-making in the medical encounter: what does it mean? (or it takes at least two to tango). Soc Sci Med 1997 Mar;44(5):681-692. [doi: 10.1016/s0277-9536(96)00221-3] [Medline: 9032835]
19. Lee K, Hoti K, Hughes JD, Emmerton L. Dr Google and the consumer: a qualitative study exploring the navigational needs and online health information-seeking behaviors of consumers with chronic health conditions. J Med Internet Res 2014 Dec 2;16(12):e262 [FREE Full text] [doi: 10.2196/jmir.3706] [Medline: 25470306]
20. Ancker JS, Witteman HO, Hafeez B, Provencher T, van de Graaf M, Wei E. The invisible work of personal health information management among people with multiple chronic conditions: qualitative interview study among patients and providers. J Med Internet Res 2015 Jun 4;17(6):e137 [FREE Full text] [doi: 10.2196/jmir.4381] [Medline: 26043709]
21. El Sherif R, Pluye P, Thoër C, Rodriguez C. Reducing negative outcomes of online consumer health information: qualitative interpretive study with clinicians, librarians, and consumers. J Med Internet Res 2018 May 4;20(5):e169 [FREE Full text] [doi: 10.2196/jmir.9326] [Medline: 29728350]
22. McMullan M. Patients using the Internet to obtain health information: how this affects the patient-health professional relationship. Patient Educ Couns 2006 Oct;63(1-2):24-28. [doi: 10.1016/j.pec.2005.10.006] [Medline: 16406474]
23. Seçkin G, Yeatts D, Hughes S, Hudson C, Bell V. Being an informed consumer of health information and assessment of electronic health literacy in a national sample of internet users: validity and reliability of the e-HLS instrument. J Med Internet Res 2016 Jul 11;18(7):e161 [FREE Full text] [doi: 10.2196/jmir.5496] [Medline: 27400726]
24. de Rosis S, Barsanti S. Patient satisfaction, e-health and the evolution of the patient-general practitioner relationship: evidence from an Italian survey. Health Policy 2016 Nov;120(11):1279-1292 [FREE Full text] [doi: 10.1016/j.healthpol.2016.09.012] [Medline: 27836231]
25. Powell J, Deetjen U. Characterizing the digital health citizen: mixed-methods study deriving a new typology. J Med Internet Res 2019 Mar 5;21(3):e11279 [FREE Full text] [doi: 10.2196/11279] [Medline: 30835238]
26. Kivits J. Informed patients and the internet: a mediated context for consultations with health professionals. J Health Psychol 2006 Mar;11(2):269-282. [doi: 10.1177/1359105306061186] [Medline: 16464924]
27. United Nations Publications. World Population Prospects: Key Findings and Advance Tables. New York, NY: United Nations; 2015.
28. Schulz PJ, Fitzpatrick MA, Hess A, Sudbury-Riley L, Hartung U. Effects of eHealth literacy on general practitioner consultations: a mediation analysis. J Med Internet Res 2017 May 16;19(5):e166 [FREE Full text] [doi: 10.2196/jmir.6317] [Medline: 28512081]
29. Sudbury-Riley L, FitzPatrick M, Schulz PJ. Exploring the measurement properties of the eHealth literacy scale (eHEALS) among baby boomers: a multinational test of measurement invariance. J Med Internet Res 2017 Feb 27;19(2):e53 [FREE Full text] [doi: 10.2196/jmir.5998] [Medline: 28242590]
30. Agarwal J, Malhotra NK, Bolton RN. A cross-national and cross-cultural approach to global market segmentation: an application using consumers' perceived service quality. J Int Mark 2010;18(3):18-40. [doi: 10.1509/jimk.18.3.18]
31. Bruning ER, Hu MY, Hao W. Cross-national segmentation: an application to the NAFTA airline passenger market. Eur J Mark 2009;43(11/12):1498-1522. [doi: 10.1108/03090560910990009]
32. Furrer O, Liu BS, Sudharshan D. The relationships between culture and service quality perceptions. J Serv Res 2000;2(4):355-371. [doi: 10.1177/109467050024004]
33. Lockshin L, Cohen E. Using product and retail choice attributes for cross‐national segmentation. Eur J Mark 2011;45(7/8):1236-1252. [doi: 10.1108/03090561111137697]
34. Shiu EC, Dawson JA. Cross-national consumer segmentation of internet shopping for Britain and Taiwan. Serv Ind J 2002;22(1):147-166. [doi: 10.1080/714005058]
35. Steenkamp JE, ter Hofstede F. International market segmentation: issues and perspectives. Int J Res Mark 2002;19(3):185-213. [doi: 10.1016/s0167-8116(02)00076-9]
36. Churchill GA. A paradigm for developing better measures of marketing constructs. J Mark Res 1979;16(1):64. [doi: 10.2307/3150876]
37. Boateng GO, Neilands TB, Frongillo EA, Melgar-Quiñonez HR, Young SL. Best practices for developing and validating scales for health, social, and behavioral research: a primer. Front Public Health 2018;6:149 [FREE Full text] [doi: 10.3389/fpubh.2018.00149] [Medline: 29942800]
38. Johnson DR, Creech JC. Ordinal measures in multiple indicator models: a simulation study of categorization error. Am Sociol Rev 1983;48(3):398-407. [doi: 10.2307/2095231]

39. Norman G. Likert scales, levels of measurement and the 'laws' of statistics. Adv in Health Sci Educ 2010;15(5):625-632. [doi: 10.1007/s10459-010-9222-y] [Medline: 20146096]
40. Sullivan GM, Artino AR. Analyzing and interpreting data from likert-type scales. J Grad Med Educ 2013 Dec;5(4):541-542 [FREE Full text] [doi: 10.4300/JGME-5-4-18] [Medline: 24454995]
41. Zumbo BD, Zimmerman DW. Is the selection of statistical methods governed by level of measurement? Can Psychol 1993;34(4):390-400. [doi: 10.1037/h0078865]
42. HealthData. 2014. Health Information National Trends Survey URL: https://healthdata.gov/dataset/health-information-national-trends-survey-hints [accessed 2019-11-08]
43. Fox S, Duggan M. Pew Research Center. Health Online 2013 URL: https://www.pewinternet.org/wp-content/uploads/sites/9/media/Files/Reports/PIP_HealthOnline.pdf [accessed 2019-11-08]
44. Hess AC, Melnyk V. Pink or blue? The impact of gender cues on brand perceptions. Eur J Mark 2016;50(9/10):1550-1574. [doi: 10.1108/EJM-11-2014-0723]
45. Barnes SJ, Bauer HH, Neumann MM, Huber F. Segmenting cyberspace: a customer typology for the internet. Eur J Mark 2007;41(1/2):71-93. [doi: 10.1108/03090560710718120]
46. Loewenthal KM. An Introduction to Psychological Tests and Scales. Second Edition. Hove: Psychology Press; 2001.
47. Fornell C, Larcker DF. Structural equation models with unobservable variables and measurement error: Algebra and Statistics. J Mark Res 1981;18(3):39-50. [doi: 10.2307/3150980]
48. Shevlin M, Miles JN. Effects of sample size, model specification and factor loadings on the GFI in confirmatory factor analysis. Pers Individ Dif 1998;25(1):85-90. [doi: 10.1016/S0191-8869(98)00055-5]
49. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct Equ Modeling 1999;6(1):1-55. [doi: 10.1080/10705519909540118]
50. Awang Z. Structural Equation Modeling using AMOS Graphic. Shah Alam: Universiti Teknologi MARA Publication Centre (UPENA); 2012.
51. MacCallum RC, Browne MW, Sugawara HM. Power analysis and determination of sample size for covariance structure modeling. Psychol Method 1996;1(2):130-149. [doi: 10.1037/1082-989X.1.2.130]
52. Hair JF, Black WC, Babin BJ, Anderson RE, Tatham RL. Multivariate Data Analysis. Volume 6. Upper Saddle River, NJ: Pearson Prentice Hall; 2006.
53. Essén A, Värlander SW, Liljedal KT. Co-production in chronic care: exploitation and empowerment. Eur J Mark 2016;50(5/6):724-751. [doi: 10.1108/EJM-02-2015-0067]
54. Wald HS, Dube CE, Anthony DC. Untangling the Web--the impact of internet use on health care and the physician-patient relationship. Patient Educ Couns 2007 Nov;68(3):218-224. [doi: 10.1016/j.pec.2007.05.016] [Medline: 17920226]
55. Ybarra ML, Suman M. Help seeking behavior and the internet: a national survey. Int J Med Inform 2006 Jan;75(1):29-41. [doi: 10.1016/j.ijmedinf.2005.07.029] [Medline: 16129659]
56. Benigeri M, Pluye P. Shortcomings of health information on the internet. Health Promot Int 2003 Dec;18(4):381-386. [doi: 10.1093/heapro/dag409] [Medline: 14695369]
57. McColl-Kennedy JR, Vargo SL, Dagger TS, Sweeney JC, Kasteren YV. Health care customer value cocreation practice styles. J Serv Res 2012;15(4):370-389. [doi: 10.1177/1094670512442806]

Abbreviations

AVE: average variance extracted
CFA: confirmatory factor analysis
CR: compositional reliability
eHealth: electronic health

Edited by G Eysenbach; submitted 02.07.19; peer-reviewed by M Tkáč, C Stiles-Shields, T Muto, JA Sim, M Laurent, L Sbaffi; comments to author 24.07.19; revised version received 12.09.19; accepted 17.10.19; published 20.11.19.

Please cite as:
FitzPatrick MA, Hess AC, Sudbury-Riley L, Schulz PJ
A Typology of Patients Based on Decision-Making Styles: Cross-Sectional Survey Study
J Med Internet Res 2019;21(11):e15332
URL: http://www.jmir.org/2019/11/e15332/
doi:10.2196/15332
PMID:31746770

Original Paper

Measurement Properties of the Online EuroQol-5D-Youth Instrument in Children and Adolescents With Type 1 Diabetes Mellitus: Questionnaire Study

Karina Mayoral[1,2,3], MPH; Luis Rajmil[4], MPH, PhD, MD; Marta Murillo[5], PhD, MD; Olatz Garin[1,3,6], MPH, PhD; Angels Pont[1,3], BSc; Jordi Alonso[1,3,6], MPH, PhD, MD; Joan Bel[5], PhD, MD; Jacobo Perez[7,8], MD; Raquel Corripio[7,8], PhD, MD; Gemma Carreras[9,10], MD; Javier Herrero[11], MD; Jose-Maria Mengibar[12], MD; Dolors Rodriguez-Arjona[4], BASoc; Ulrike Ravens-Sieberer[13], MPH, PhD, MD; Hein Raat[14], MBA, PhD, MD; Vicky Serra-Sutton[3,4], BASoc, PhD; Montse Ferrer[1,2,3], MPH, PhD, MD

[1]Health Services Research Group, Hospital del Mar Medical Research Institute, Barcelona, Spain

[2]Department of Paediatrics, Obstetrics and Gynaecology and Preventive Medicine, Universitat Autònoma de Barcelona, Barcelona, Spain

[3]Centro de Investigación Biomédica en Red de Epidemiología y Salud Pública, Madrid, Spain

[4]Agency for Health Quality & Assessment of Catalonia, Barcelona, Spain

[5]Pediatric Service, Department of Pediatric Endocrinology, University Hospital Germans Trias i Pujol, Barcelona, Spain

[6]Experimental and Health Sciences, Pompeu Fabra University, Barcelona, Spain

[7]Department of Pediatric Endocrine, Hospital of Sabadell, Corporació Sanitària Parc Taulí, Sabadell, Spain

[8]University Institute Parc Taulí, Universitat Autònoma de Barcelona, Sabadell, Spain

[9]Pediatric Endocrinology, Hospital de la Santa Creu i Sant Pau, Barcelona, Spain

[10]Medicine Department, Universitat Autònoma de Barcelona, Barcelona, Spain

[11]Corporació de Salut del Maresme i la Selva, Hospital de Calella, Calella, Spain

[12]Corporació de Salut del Maresme i la Selva, Hospital de Blanes, Blanes, Spain

[13]Department of Child and Adolescent Psychiatry, Psychotherapy, and Psychosomatics, University Medical Center Hamburg - Eppendorf, Hamburg, Germany

[14]Department of Public Health, Erasmus University Medical Center Rotterdam, Rotterdam, Netherlands

Corresponding Author:
Montse Ferrer, MPH, PhD, MD
Health Services Research Group
Hospital del Mar Medical Research Institute
Doctor Aiguader, 88
Office 144
Barcelona, 08003
Spain
Phone: 34 933160763
Email: mferrer@imim.es

Abstract

Background: The lack of continuity between health-related quality of life (HRQoL) instruments designed for children and adults hinders change analysis with a life course approach. To resolve this gap, EuroQol (EQ) developed the EQ-5D-Youth (EQ-5D-Y), derived from the EQ-5D for adults. Few studies have assessed the metric properties of EQ-5D-Y in children with specific chronic conditions, and none have done so for children with type I diabetes mellitus (T1DM).

Objective: This study aimed to evaluate the acceptability, validity, reliability, and responsiveness of the EQ-5D-Y in children and adolescents with T1DM, when administered online.

Methods: Participants with T1DM were consecutively recruited from July to December 2014, from a list of potential candidates aged 8-19 years, who attended outpatient pediatric endocrinology units. Before every quarterly routine visit, participants received an email/telephone reminder to complete the online version of two generic HRQoL questionnaires: EQ-5D-Y and KIDSCREEN-27. The EQ-5D-Y measures five dimensions, from which an equally weighted summary score was constructed (range: 0-100).

Completion rate and distribution statistics were calculated. Construct validity was evaluated through known group comparisons based on general health, acute diabetic decompensations, mental health, family function, and a multitrait, multimethod matrix between EQ-5D-Y and KIDSCREEN by using Spearman correlations. Construct validity hypotheses were stated a priori. Reliability was assessed with the intraclass correlation coefficient and responsiveness by testing changes over time and calculating the effect size. Reliability and responsiveness were tested among the stable and improved subsamples defined by a KIDSCREEN-10 index change of <4.5 points or ≥4.5 points, respectively, from the first to the fourth visit.

Results: Of the 136 participants, 119 (87.5%) responded to the EQ-5D-Y at the last visit. The dimensions that showed higher percentages of participants with problems were "having pain/discomfort" (34.6%) and "worried/sad/unhappy" (28.7%). The mean (SD) of the EQ-5D-Y summary score was 8.5 (10.9), with ceiling and floor effects of 50.7% and 0%, respectively. Statistically significant HRQoL differences between groups defined by their general health (excellent/very good and good/regular/bad) and mental health (Strengths and Difficulties Questionnaire score ≤15 and >16, respectively) were found in three EQ-5D-Y dimensions ("doing usual activities," "having pain/discomfort," and "feeling worried/sad/unhappy"), summary score (effect size for general health and mental health groups=0.7 and 1.5, respectively), and KIDSCREEN-10 index (effect size for general health and mental health groups=0.6 and 0.9, respectively). Significant differences in the EQ-5D-Y dimensions were also found according to acute diabetic decompensations in "looking after myself" (P=.005) and according to family function in "having pain/discomfort" (P=.03). Results of the multitrait, multimethod matrix confirmed three of the four relationships hypothesized as substantial (0.21, 0.58, 0.50, and 0.46). The EQ-5D-Y summary score presented an intraclass correlation coefficient of 0.83. Statistically significant change between visits was observed in the improved subsample, with an effect size of 0.7 (P<.001).

Conclusions: These results support the use of the EQ-5D-Y administered online as an acceptable, valid, reliable, and responsive instrument for evaluating HRQoL in children and adolescents with T1DM.

(J Med Internet Res 2019;21(11):e14947) doi:10.2196/14947

KEYWORDS

health-related quality of life; type 1 diabetes; EQ-5D-Y; EuroQol; validity

Introduction

Type 1 diabetes mellitus (T1DM) is one of the most common chronic childhood illnesses, affecting approximately 1 in every 400-600 children and adolescents [1]. Treatment includes a multifaceted regimen with daily subcutaneous insulin injections (or insulin infusion), blood glucose monitoring, carbohydrate counting, dietary plan, and physical activity [2]. Due to its complexity, T1DM management can be overwhelming even for the most competent patient [3]. Children might feel different from their peers because of their need for disruptive self-care activities [4] and the impact of T1DM on psychological aspects [5]. Health-related quality of life (HRQoL) captures the individual's health perception, adding the patients' perspective into T1DM clinical monitoring and research.

The lack of correspondence and continuity between HRQoL instruments designed for children and those designed for adults hinders the analysis of HRQoL changes using a life course approach. To solve this gap, in 2006, the EuroQol Group developed the EQ-5D-Youth (EQ-5D-Y) as a version of the EQ-5D for adults. The EQ-5D-Y was designed by adapting the EQ-5D dimensions to the requirements of HRQoL measurement in children and adolescents aged 8 years or older [6,7]. The main rationale of the new version was to enable young individuals to self-report their health [6,8]. A EQ-5D-Y proxy version was also developed [9] for children below 8 years of age. Given its generic and econometric nature, it allows comparison between different populations and conditions, and cost-utility analysis for economic evaluation. Another econometric instrument, the Health Utility Index [10], has a self-administered version for adults and adolescents and a proxy version for children (5-12 years of age), but its administration burden is substantially greater.

The EQ-5D-Y is a very short (2-3 minutes) and simple instrument to fill out. There are several studies assessing the metric properties of the EQ-5D-Y in the general population [11,12] or in school environments [13-16]. A Swedish study [12] suggested that the EQ-5D-Y was comprehensible, acceptable, and feasible for self-completion. A multicountry study supported its feasibility, reliability, and validity [11]. Furthermore, a Spanish study [16] compared the paper and Web-based versions of EQ-5D-Y, showing acceptable levels of agreement.

Only a few studies have assessed the metric properties of EQ-5D-Y in children with specific chronic conditions. A multicenter study among patients with cystic fibrosis [17] concluded that the EQ-5D-Y can be considered a valid instrument, which reflects the differences in health according to the progression of this life-long chronic disease. A comparison between the EQ-5D-Y and the Pediatric Asthma Quality of Life Questionnaire [18] supported EQ-5D-Y's convergent validity for asthmatic children and adolescents. In addition, a study in youth with chronic kidney disease provided evidence of the EQ-5D-Y validity, as it discriminated among groups that differed with regard to the disease-related clinical burden [19]. Finally, we have only identified one descriptive study of HRQoL in children with T1DM [20] measured with the EQ-5D-Y, but without assessing its metric properties. Our aim was to evaluate acceptability, validity, reliability and responsiveness of the EQ-5D-Y administered online in children and adolescents with T1DM.

Data was obtained from a clinical trial on children and adolescents with T1DM, designed to evaluate the impact of routine use of HRQoL assessment in clinical practice by using the KIDSCREEN-27 as the primary outcome. Previous publications of this trial reported that the baseline HRQoL of children and adolescents with T1DM [21] was similar to that of the European population of the same age, but with slightly worse physical well-being, and that routine assessment and face-to-face patient-physician discussion of HRQoL results improved after a year of follow-up, especially psychological well-being and school environment [22].

Methods

Participants and Study Design

Subjects were consecutively recruited from July to December 2014, from a list of 205 potential candidates with T1DM, aged between 8 and 19 years, attending outpatient pediatric endocrinology units of 5 hospitals. Exclusion criteria were as follows: T1DM diagnosed within the last 6 months and presence of cognitive problems that prevented comprehension of the questionnaires. Seven pediatricians were randomly assigned to either the intervention or control group (four and three pediatricians, respectively).

Families of children that fulfilled the inclusion criteria were informed about the project, and those who agreed to participate received a reminder by email and then by telephone, if necessary, before every quarterly routine visit to complete the online questionnaires within the 48 hours prior to the visit (for children with limited access to internet, a laptop was available at the doctor's office).

The intervention involved a discussion of the HRQoL scores between the doctor and the participant at each routine visit. The KIDSCREEN-27 and EQ-5D-Y results were displayed with different colors, similar to a traffic light, marking green for good, yellow for medium, and red for poor outcomes. Evolutionary results were displayed in the second, third, and fourth visits, to show a comparison with the previous results. At each visit, pediatricians invited the patients to comment and discuss the results, identifying dimensions with low value scores, exploring possible solutions and actions, and coming back to these actions over time with advice. Before starting the study, the pediatricians in this group had received standardized training in the use and interpretation of HRQoL questionnaires.

The control group also completed the questionnaires online. However, the questionnaires completed by the control group 48 h before each visit were about physical activity and diet, and the HRQoL questionnaires were completed only before the first and fourth visits. During the consultation, the patients received usual care without commenting on the results of the questionnaires (physicians did not receive any instruction on clinical management).

Study Variables

Information collected from clinical records included age, gender, weight, height, and glycated hemoglobin (HbA_{1c}) levels. Acute diabetic decompensations during the previous 3 months were registered: hypoglycemia (<60 mg/dL) with decreased level of consciousness, requiring glucagon or the help of others to recover, and hyperglycemia (>400-450 mg/dL), which required intervention of professionals or presented with diabetic ketoacidosis.

HRQoL evaluation consisted of the administration of the Spanish online version of the KIDSCREEN-27 and the EQ-5D-Youth (EQ-5D-Y). The descriptive system of the EQ-5D-Y [6] consists of five dimensions ("mobility," "looking after myself," "doing usual activities," "having pain/discomfort," and "feeling worried/sad/unhappy") with three-level Likert scale responses (no problems, moderate problems, and serious problems). It includes also a visual analogue scale (EQ-VAS) on the general health status rated from 0 (worst health status) to 100 (best health status possible). The time frame for both dimensions and EQ-VAS is "today." As no preference weights were available for the EQ-5D-Y to estimate utilities [23], an equally weighted summary score of the five dimensions was calculated [24-26], ranging from 0 (no problem in any dimension) to 100 (serious problems in all dimensions).

The KIDSCREEN-27 [27] contains 5 dimensions: physical well-being (5 items), psychological well-being (7 items), autonomy and relationships with parents (7 items), social support and relationship with friends (4 items), and school environment (4 items). Responses were categorized into five-option Likert scales that assess the frequency or intensity of the attribute, with a recall period of 1 week in most questions. Scores were calculated following developers' recommendations [27] for the KIDSCREEN-27 dimensions and for the KIDSCREEN-10 index, constructed with 10 items that sufficiently represent the longer KIDSCREEN profiles. The KIDSCREEN scores are standardized to a mean of 50 and a SD of 10, from a reference sample of 22,000 European children and adolescents [28].

In addition to HRQoL, children's mental health status and family function were assessed. The Strengths and Difficulties Questionnaire (SDQ) [29] consists of 25 items measuring a range of mental health symptoms including conduct problems, hyperactivity-inattention, emotional symptoms and peer problems, and prosocial behaviors. All items are scored on a 3-point scale (0=not true, 1=somewhat true, and 2=certainly true). Items are summed to give a total difficulties score ranging from 0 (no problems) to 40 (maximum problems). The Spanish version of the SDQ has shown to be reliable and valid [30]. The family function questionnaire [31,32] was designed to measure the patient's satisfaction with the support from the family through five items: Adaptation, Partnership, Growth, Affection, and Resolve (APGAR). These items have three response options (from almost always to hardly ever), and the total score ranges from 0 to 10 (low to high satisfaction).

Ethics Considerations

The study was approved by the ethics committee of participant hospitals in accordance with national and international guidelines (code of ethics, Helsinki Declaration) as well as legislation on data confidentiality (Organic Law 15/1999 of December 13 Data Protection character staff). Signed consent to participate was requested from parents and children over 12

years. The collection and transfer of data were carried out according to strict security and data encryption.

Analytical Strategy

Considering the 136 patients included in the T1DM clinical trial, this sample size gave a statistical power of 0.80 to detect a moderate-to-large difference of 0.65 SDs in the EQ-5D-Y summary score between unequally distributed known groups (80% and 20% of participants in each category) using a two-sided test with type I error of 5%. Statistical power was calculated using published formulas [33].

Characteristics of the sample were described by calculating percentages, or means and SDs, according to the type of variable. To evaluate the acceptability of the EQ-5D-Y, we calculated the completion rate and the distribution of the response options in each dimension. Distribution of the EQ-5D-Y summary score, EQ-VAS, and the KIDSCREEN-10 index was described by calculating the observed range, floor, and ceiling effects (proportion of participants with the worst and best possible score, respectively) and statistics of central tendency and dispersion.

Construct validity was assessed by applying two different approaches: (1) comparison of known groups and (2) the multitrait, multimethod matrix with data from the first visit. Known groups were defined according to the general health (through a question from KIDSCREEN-27), dichotomized as excellent/very good and good/regular/bad; acute diabetic decompensations during the previous 3 months; mental health, dichotomized by the total SDQ score cutoff of 15; and family, classified as dysfunctional (total APGAR score≤6) or functional (total APGAR score=7-10) [32]. To assess differences between known groups, a Chi-square test was used for proportions of participants with problems, Wilcoxon nonparametric test was used for the EQ-5D-Y summary score and EQ-VAS, and unpaired *t*-test was used for KIDSCREEN scores. The magnitude of the differences between groups was assessed by the Cohen effect size (difference of mean/pooled SD) [34]. General guidelines define an effect size of 0.2 as small, 0.5 as moderate, and 0.8 as large [35]. The hypotheses raised a priori, based on available evidence, were higher EQ-5D-Y summary score (worse HRQoL) in children reporting worse general health [11], with previous acute diabetic decompensations [36,37], worse mental health [11,38], and with a dysfunctional family [39,40].

The multitrait, multimethod matrix between the EQ-5D-Y and the KIDSCREEN was constructed using Spearman correlations. The logical relationship between two instruments can be categorized as convergent and discriminant. For the convergent validity [7] (different instruments measuring a similar concept), we hypothesized substantial correlations between “mobility” (EQ-5D-Y) and “physical well-being” (KIDSCREEN-27), “feeling worried/sad/unhappy” (EQ-5D-Y) and “psychological wellbeing” (KIDSCREEN-27), and the KIDSCREEN-10 index with the EQ-5D-Y summary score and EQ-VAS. In contrast, for the discriminant validity [7] (different instruments measuring different traits or constructs), we hypothesized low correlations between the EQ-5D-Y dimension of “looking after myself” and the following three KIDSCREEN-27 dimensions: “autonomy and relationships with parents,” “social support and relationship with friends,” and “school environment.” The strength of Spearman correlations was defined [41] as insignificant (≤0.30), moderate (0.31-0.44), or substantial (0.45-0.60).

The reliability was assessed in terms of reproducibility (stability of an instrument over time) in a subsample of stable participants—those in the control group with an absolute change <4.5 points in the KIDSCREEN-10 index from the first to the fourth visit, which corresponds to a small magnitude of change (effect size<0.45). To measure the agreement in EQ-5D-Y summary score and EQ-VAS between both administrations, the intraclass correlation coefficient (ICC) was calculated. To assess responsiveness, we classified participants experiencing “improvement” as those with a change in KIDSCREEN-10 index ≥4.5 points (moderate effect size) between the first and the fourth visits. Differences for the EQ-5D-Y dimensions were compared using the McNemar paired test, while differences between the EQ-5D-Y summary score and EQ-VAS were tested with the Wilcoxon paired test. The magnitude of change was also measured by the effect size coefficient (mean of change/SD of change). Program STATA.14 software (StateCorp LP, College Station, Texas) was used in the analysis.

Results

Sample Characteristics

Of 205 potential candidates with T1DM, 61 were not included due to change of address, transfer to the adult unit, or no attendance at the follow-up visits; in addition, 8 subjects rejected participation. Finally, 136 participants were included in the study. Table 1 shows the characteristics of the sample. Mean age was 14 years, around half of the participants were girls, 71 (52.2%) had HbA_{1c} levels ≤7.5% (58 mmol/mL), 123 (90.4%) did not present acute diabetic decompensations during the previous 3 months, and the majority (n=95) rated their health as good or very good.

Table 1. Demographic and clinical characteristics of the participants (N=136).

Characteristic	Value
Age (years)[a], mean (SD)	13.99 (0.2)
8-12, n (%)	45 (33.3)
13-15, n (%)	54 (40.0)
16-18, n (%)	35 (25.9)
>18, n (%)	1 (0.7)
Sex, n (%)	
Girls	72 (52.9)
Boys	64 (47.1)
Metabolic control, HbA_{1c}[b], mean (SD)	7.7 (1.3)
>7.5, n (%)	65 (47.8)
≤7.5, n (%)	71 (52.2)
Acute diabetic decompensations (previous 3 months), n (%)	
Yes	13 (9.6)
No	123 (90.4)
General Health Question (KIDSCREEN-27), n (%)	
Excellent	19 (13.9)
Very good	44 (32.4)
Good	54 (39.7)
Regular	17 (12.5)
Bad	2 (1.5)
Mental health (total SDQ[c] score), mean (SD)	10.64 (5.3)
0-15, n (%)	113 (83.1)
16-40, n (%)	23 (16.9)
Family function (APGAR[d]), n (%)	
Dysfunctional	52 (38.2)
Functional	84 (61.8)

[a]The age of one participant is missing (N=135).

[b]HbA_{1c}: glycated hemoglobin.

[c]SDQ: Strengths and Difficulties Questionnaire.

[d]APGAR: Adaptation, Partnership, Growth, Affection, and Resolve.

Distribution of Dimensions and Indices of Health-Related Quality of Life

Children and adolescents with T1DM reported relatively fewer health problems in the EQ-5D-Y dimensions (Table 2), especially for "mobility" (n=3, 2.2%) and "looking after myself" (n=4, 2.9%). Dimensions showing higher percentages of participants with problems were "having pain or discomfort" (n=47, 34.6%) and "feeling worried/sad/unhappy" (n=39, 28.6%). Means and 95% CI of KIDSCREEN-27 dimensions (Table 3) were below 50 (European sample mean [28]) for "physical well-being" (46.2, 95% CI 44.7-47.7) and over 50 for "school environment" (52.0, 95% CI 50.4-53.6) and "social support/relationship with friends" (53.6, 95% CI 52.1-55.1). As shown in Table 4, the mean (SD) of the EQ-5D-Y summary score was 8.5 (10.9), with ceiling and floor effects of 50.7% and 0%, respectively. The EQ-VAS ranged in our sample from 31 to 100, with a mean (SD) of 80.4 (14.7). The KIDSCREEN index ranged from 40.0 to 83.8, and the mean (SD) was 49.7 (8.3).

Table 2. Distribution of the EuroQol-5D-Youth (EQ-5D-Y) dimensions.

EQ-5D-Y dimensions	No problems, n (%)	Some problems, n (%)	A lot of problems, n (%)
Mobility	133 (97.8)	3 (2.2)	0 (0.0)
Looking after myself	132 (97.1)	4 (2.9)	0 (0.0)
Doing usual activities	119 (87.5)	17 (12.5)	0 (0.0)
Having pain/discomfort	89 (65.4)	45 (33.1)	2 (1.5)
Feeling worried/sad/unhappy	97 (71.3)	35 (25.7)	4 (2.9)

Table 3. Distribution of KIDSCREEN-27 dimensions.

KIDSCREEN-27 dimensions	Mean (SD), 95% CI	Median
Physical well-being	46.2 (9.0), 44.7-47.7	44.7
Autonomy/relationships with parents	50.5 (8.2), 49.1-51.9	49.5
School environment	52.0 (9.3), 50.4-53.6	51.1
Social support and relationship with friends	53.6 (9.1), 52.1-55.1	53.3
Psychological well-being	49.7 (9.7), 48.1-51.3	48.5

Table 4. Distribution of the EuroQol-5D-Youth (EQ-5D-Y) summary score, EuroQol-Visual Analog Scale (EQ-VAS), and KIDSCREEN-10 index.

Distribution of scores	EQ-5D-Y summary score	EQ-VAS	KIDSCREEN-10 index
Theoretical range	0-100	0-100	–3.54 to 83,81
Observed range	0-40	31-100	40.0-83.81
Proportion of participants with the best health (%)	50.7	8.1	0.7
Proportion of participants with the worst health (%)	0	0	0
Mean (SD)	8.5 (10.9)	80.4 (14.7)	49.7 (8.3)

Construct Validity of the EuroQol-5D-Youth Online

Table 5 shows statistically significant HRQoL differences between groups defined by their general health in three of five EQ-5D-Y dimensions as well as EQ-5D-Y summary score, EQ-VAS, and KIDSCREEN-10 index (P<.001, P<.001, and P=.02, respectively) with moderate-large effect sizes (0.7, 1.0, and 0.6, respectively). Similar significant HRQoL differences were found according to mental health with large effect sizes, except for the EQ-VAS. The only significant difference between those with and without acute diabetic decompensations was found in the “looking after myself” dimension (P=.005). Finally, HRQoL differences between functional and dysfunctional families were statistically significant for the “having pain and discomfort” EQ-5D-Y dimension (P=.03) and KIDSCREEN-10 index (P<.001).

Table 5. Comparison of health-related quality of life between known groups measured at baseline.

Variables	EQ-5D-Y[a] dimensions					EQ-5D-Y summary score, mean (SD), median	EQ-VAS[b], mean (SD), median	KIDSCREEN-10 index, mean (SD), median
	M[c]	LAM[d]	DUA[e]	HP/D[f]	W/S/U[g]			
General health question								
Excellent/very good	0.0%	3.2%	4.8%	22.2%	19.1%	5.08 (8.4), 0	87.6 (10.5), 90	52.24 (6.9), 53.1
Good/regular/bad	4.1%	2.7%	19.2%	45.2%	37.0%	11.51 (12.1), 10	74.1 (15.0), 75	47.47 (8.9), 45.7
P value	.25	1.00	.02[h]	<.01[h]	.02[h]	<.001[h]	<.001[h]	.02[h]
Effect size (95% CI)	N/A[i]	N/A	N/A	N/A	N/A	0.7 (0.39-1.09)	1.0 (0.66-1.38)	0.6 (0.25-0.94)
Acute diabetic decompensations								
Yes	7.7%	15.4%	23.1%	38.5%	46.2%	13.85 (13.2), 10.0	74.9 (18.0), 75	50.21 (9.8), 46.9
No	1.6%	1.6%	11.4%	34.2%	26.8%	7.97 (10.6), 0	80.9 (14.3), 81	49.63 (8.2), 48.3
P value	.26	.005[h]	.23	.76	.14	.08	.21	.81
Effect size (95% CI)	N/A	N/A	N/A	N/A	N/A	0.5 (–0.04 to 1.12)	0.4 (–0.17 to 0.99)	0.07 (–0.5 to 0.65)
Mental health (SDQ[j])								
Score 0-15	0.9%	2.7%	8.9%	28.3%	20.4%	6.11 (8.4), 0	81.3 (14.8), 81	50.84 (8.3), 49.8
Score 16-40	8.7%	4.4%	30.4%	65.2%	69.6%	20.43 (14.3), 20	75.7 (13.5), 75	44.01 (5.6), 43.4
P value	.07	.53	.004[h]	.001[h]	<.001[h]	<.001[h]	.06	<.001[h]
Effect size (95% CI)	N/A	N/A	N/A	N/A	N/A	1.5 (1.01-1.98)	0.4 (–0.07 to 0.83)	0.9 (0.40-1.32)
Family function (APGAR [k])								
Score≤6	1.2%	4.8%	10.7%	27.4%	25.0%	7.02 (9.9), 0	81.6 (15.1), 85	51.46 (8.6), 50.6
Score 7-10	3.9%	0.0%	15.4%	46.2%	34.6%	10.96 (12.3), 10	78.3 (13.9), 80	46.81 (6.9), 46.9
P value	.56	.30	.42	.03[h]	.23	.06	.12	<.001[h]
Effect size (95% CI)	N/A	N/A	N/A	N/A	N/A	0.4 (0.01-0.71)	0.2 (–0.12 to 0.58)	0.6 (0.22-0.93)

[a]EQ-5D-Y: EuroQol-5D-Youth.

[b]EQ-VAS: EuroQol-Visual Analog Scale.

[c]M: mobility.

[d]LAM: looking after myself.

[e]DUA: doing usual activities.

[f]HP/D: having pain/discomfort.

[g]W/S/U: feeling worried/sad/unhappy.

[h]Statistically significant difference.

[i]N/A: not applicable.

[j]SDQ: Strengths and Difficulties Questionnaire.

[k]APGAR: Adaptation, Partnership, Growth, Affection, and Resolve.

Table 6 presents the multitrait, multimethod matrix between EQ-5D-Y and KIDSCREEN. Regarding convergent validity, of the four correlations previously hypothesized as substantial, subscript a), three obtained coefficient values of 0.45 or greater: 0.58 between EQ-5D-Y "feeling worried/sad/unhappy" and KIDSCREEN "psychological wellbeing," 0.50 between the EQ-5D-Y summary score and KIDSCREEN-10 index, and 0.46 between the EQ-VAS and KIDSCREEN-10 index. In contrast, the correlation between the mobility dimension of the EQ-5D-Y and the physical well-being dimension of the KIDSCREEN-27 was 0.21. For discriminant validity, the three relationships hypothesized as insignificant (subscript b) ranged from 0.02 to 0.06, as expected.

Table 6. Multitrait, multimethod matrix between the EuroQol-5D-Youth (EQ-5D-Y) and the KIDSCREEN in children and adolescents with type 1 diabetes mellitus at baseline evaluation.

EQ-5D-Y	KIDSCREEN-27					KIDSCREEN-10 index
	Physical well-being	Psychological well-being	Autonomy and relationships with parents	Social support and relationship with friends	School environment	
Mobility	0.21[a]	0.19	0.02	0.06	0.12	0.16
Looking after myself	0.02	0.07	0.06[b]	0.02[b]	0.06[b]	0.03
Doing usual activities	0.17	0.29	0.08	0.14	0.25	0.25
Having pain/discomfort	0.27	0.31	0.23	0.13	0.17	0.32
Feeling worried/sad/unhappy	0.30	0.58[a]	0.22	0.36	0.40	0.56
EQ-5D-Y summary score	0.23	0.10	0.01	0.18	0.00	0.50[a]
EuroQol-Visual Analog Scale	0.38	0.31	0.21	0.14	0.35	0.46[a]

[a]Correlations hypothesized as substantial (0.45-0.60).

[b]Correlations hypothesized as insignificant (≤0.30).

Reproducibility and Responsiveness

Table 7 shows EQ-5D-Y test-retest results in the "stable" subsample between the first and the fourth visit. The ICC was 0.83 and 0.74 for the EQ-5D-Y summary score and EQ-VAS, demonstrating stability over time for this subsample. Regarding responsiveness, this "improvement" subsample presented significant changes in three dimensions, in the EQ-5D-Y summary score, and EQ-VAS between the first visit and the fourth visit. The effect sizes were 0.70 and 0.74, indicating an improvement of moderate-to-large magnitude.

Table 7. Reproducibility and responsiveness of EuroQol-5D-Youth (EQ-5D-Y): problems reported by each dimension, summary score, and EuroQol-Visual Analog Scale (EQ-VAS) at the first and fourth visits.

EQ-5D-Y	Test-retest reproducibility among stable participants (n=18)			Responsiveness among participants with improvement (n=58)		
	First visit	Fourth visit	*P* value	First visit	Fourth visit	*P* value
EQ-5D-Y dimensions						
Mobility, n (%)	0 (0.0)	1 (5.6)	—[a]	3 (5.2)	1 (1.7)	.63
Looking after myself, n (%)	1 (5.6)	0 (0.0)	—	2 (3.5)	1 (1.7)	>.99
Doing usual activities, n (%)	1 (5.6)	1 (5.6)	.06	9 (15.5)	2 (3.5)	.02[b]
Having pain/discomfort, n (%)	4 (22.2)	4 (22.2)	.13	27 (46.6)	11 (18.9)	<.001[b]
Feeling worried/sad/unhappy, n (%)	6 (33.3)	5 (27.8)	.14	23 (39.7)	10 (17.2)	<.001[b]
EQ-5D-Y summary score			.71			<.001[b]
Mean (SD)	6.67 (9.1)	6.11 (7.8)		11.90 (12.6)	4.31(7.0)	
Effect size (95% CI)	N/A[c]	0.07 (–0.61 to 0.74)		N/A	0.70 (0.31 to 1.02)	
ICC[d] (95% CI)	N/A	0.83 (0.55 to 0.94)		N/A	N/A	
EQ-VAS			.15			<.001[b]
Mean (SD)	82 (13.5)	86 (9.1)		76.0 (16.6)	86.9 (13.0)	
Effect size (95% CI)	N/A	0.35 (–0.34 to 1.03)		N/A	0.74 (0.36 to 1.12)	
ICC (95% CI)	N/A	0.74 (0.31 to 0.90)		N/A	N/A	

[a]Not available.

[b]Statistically significant difference.

[c]N/A: not applicable.

[d]ICC: intraclass correlation coefficient.

Discussion

To the best of our knowledge, this is the first study evaluating metric properties of the new EQ-5D-Y in T1DM children and adolescents. We found this generic preference-based instrument to be acceptable for children and adolescents with T1DM and easy to administer online, but with a high ceiling effect (50.7% of participants reported no problem in any dimension). The EQ-5D-Y showed good validity, considering the results obtained in the known groups' analysis based on general health, acute diabetic decompensations, mental health, and family function, and those obtained from the multitrait, multimethod matrix with KIDSCREEN. Test-retest reproducibility was high, indicating good reliability, and the moderate-large change observed between the first and fourth visits among the selected subsample of "improvement" supports its responsiveness.

Of the 136 participants in the T1DM study, all responded to the EQ-5D-Y at the first visit and 119 (87.5%) responded at the fourth visit, which is similar to the response rates reported by studies carried out in the clinical (96% [18]) and school (77% [16]) settings. It is important to remark that the participants answering the EQ-5D-Y entirely completed the instrument, including the five dimensions, and the EQ-VAS. These rates of completion support the acceptability of the EQ-5D-Y among children and adolescents with T1DM when administered online. Taking into account the quick expansion of online recreational habits among the new generations born in the digital era, acceptability could be even higher with the incorporation of other devices such as apps for smartphones or tablets. However, completion rates in routine clinical practice could be lower than under clinical trial conditions, which indicates the need for closer monitorization.

Ceiling effect in the sample of children and adolescents with T1DM was high for the EQ-5D-Y summary score (50.7%) and for the three out of its five dimensions, which exceeded 85% of participants reporting no problem ("mobility," "looking after myself," and "doing usual activities"). A European multicountry study of EQ-5D-Y metric properties [11] reported a similar ceiling effect for these dimensions, which was up to 80% in children with chronic conditions. Half of the children and adolescents with no problem in any dimension (value=0 in EQ-5D-Y summary score) could be considered to show a really high ceiling effect, taking into account the established recommendations of 15% for HRQoL scores [42]. However, this is a general standard, as there are none specific for children. This high ceiling effect could be considered coherent with the relative low impact of diabetes on HRQoL for most children and adolescents, partially explained by their capacity of adaptation to the demanding tasks related to care [43]. The main disadvantage of this high ceiling effect is that it avoids the detection of improvement in children and adolescents who reported the highest level of health, which could affect the instrument's responsiveness.

On the other hand, our study showed a substantial proportion of children and adolescents with T1DM reporting problems in the dimensions of "having pain or discomfort" (n=47, 34.6%) and "feeling worried/sad/unhappy" (n=43, 28.7%). These results are consistent with a previous study [44] showing this latter EQ-5D-Y dimension as the most affected in children and adolescents with T1DM. They are also consistent with studies using the 21-item Beck Depression Inventory and the Child Health Questionnaire PF-50 [45,46], which highlighted the impact of T1DM on the psychological dimension, especially due to the lack of autonomy, worrying for future chronic complications and the self-care routines that these children and adolescents have to experience.

As hypothesized, the EQ-5D-Y presented a good discrimination capacity between groups defined by the general health question and SDQ score in the summary score and in three dimensions: "doing usual activities," "having pain/discomfort," and "feeling worried/sad/unhappy." These dimensions presented the highest correlations with general and mental health in previous studies of children and adolescents with T1DM [41] or chronic conditions [10]. The magnitude of the difference between groups defined by these constructs was larger when HRQoL was measured with the EQ-5D-Y summary score (effect size for general health and mental health groups=0.7 and 1.5, respectively) than with the KIDSCREEN-10 index (effect size for general health and mental health groups=0.6 and 0.9, respectively).

It is important to highlight the problems reported in the "looking after myself" dimension among those participants who had previously experienced acute diabetic decompensations, which is consistent with previous studies comparing good and bad metabolic control [37,47]. These participants showed differences in the physical well-being and health perception dimensions, measured with KINDL-R [37] and the Diabetes Quality of Life questionnaire [42]. A priori hypotheses on the relationship between family function and HRQoL were based on studies that reported an increased risk for depressive symptoms in families with a disadvantaged socioeconomic status [39,48]. Our results, showing problems of "having pain and discomfort" more frequently and poorer HRQoL with KIDSCREEN-10 index in children and adolescents with dysfunctional families, are in line with the social component of health [39,48].

Correlations between EQ-5D-Y and KIDSCREEN showed the expected pattern for convergent and divergent validity, which were very similar to those obtained in the abovementioned European multicountry study [11]. The dimensions reflecting the emotional aspects of EQ-5D-Y ("feeling worried/sad/unhappy") and KIDSCREEN-27 ("psychological well-being"), in addition to the EQ-5D-Y summary score and EQ-VAS, presented a substantial correlation with the KIDSCREEN-10 index, which provides a summary of the HRQoL. The EQ-5D-Y "mobility" dimension failed to display convergent validity with the KIDSCREEN dimension "physical wellbeing," similar to a previous study [11] (r=0.10-0.25). Considering the content of the instruments, it can be argued that KIDSCREEN is more focused on the energy level and less focused on the physical functioning, which is the case of the EQ-5D-Y.

Our study shows good reproducibility of the questionnaire according to the established standards [7] for reliability coefficients: 0.7 and 0.9 for group and individual measurements,

respectively. The ICC of 0.83 and 0.74 obtained for the EQ-5D-Y summary score and EQ-VAS allows recommending its use. Nine months between the first and the fourth visit is too long a period to evaluate reproducibility, since HRQoL could change considerably during this time. This long period explains the low number of children and adolescents in the stable subsample (n=18). Previous EQ-5D-Y studies used the kappa coefficient to evaluate test-retest reporting from good to fair agreement, according to dimensions [11,15].

The EQ-5D-Y detected differences in the selected subsample of children and adolescents with T1DM classified as experiencing "improvement" between the first and the fourth visit, showing the capacity of this instrument to detect change. It is important to highlight that no other studies testing the responsiveness of the EQ-5D-Y have been found. A systematic review assessing metric properties of diabetes-specific HRQoL instruments also highlighted that the lack of testing for responsiveness is a major shortcoming [49]. The EQ-5D-Y summary score indicates a moderate-large improvement over time, a magnitude which can be considered reasonable when accounting for the length of the follow-up of participants (9 months), the nature of the intervention, and the course of the T1DM.

The T1DM clinical trial provides a comprehensive and complete database of EQ-5D-Y and disease-related variables to allow assessment of metric properties including validity, reliability, and responsiveness. However, some limitations need to be addressed. First, the assessment of the acceptability is not complete: Other indicators such as administration time needed to complete the questionnaire and patient views about the instrument should be included in further studies. Second, since the EuroQol group has not developed the value set for the EQ-5D-Y yet [23], the unweighted summary score was calculated as was done in previous studies [24-26]. However, this summary score did not allow cost-utility analysis. Third, the EQ-5D-Y does not cover social dimension, a key aspect of children's HRQoL, but it presents an acceptable capacity of discrimination between functional and dysfunctional families in our study. There are other generic pediatric instruments that cover social dimensions, such as KIDSCREEN [50] ("autonomy and relationships with parents," "social support," "relationship with friends," and "school environment"), PedsQol [51] ("social functioning" and "school functioning"), or the Child Health Questionnaire [52] ("role/social emotional and behavioral functioning"). Fourth, this is a secondary analysis of a study designed for purposes other than the evaluation of the metric properties of EQ-5D-Y. For example, as commented above, the 9-month period between test and retest is too long for measuring reproducibility, and no large change is expected with the intervention applied. Furthermore, results on stability should be interpreted with caution due to the small sample size. Finally, regarding external validity, the results of this study should not be generalized to other chronic conditions.

Despite the limitations discussed above, our results provide considerable evidence supporting the appropriate metric properties of the EQ-5D-Y in patients with T1DM. In conclusion, these findings suggest that the EQ-5D-Y administered online is an acceptable, valid, reliable, and responsive instrument for evaluating HRQoL in children and adolescents with this chronic condition. Since it is a preference-based health status measure, it will allow calculating quality-adjusted life-years (combining both the quantity and quality of life) for economic evaluations, once social preferences specifically for children become available. Given its short and easy administration, the EQ-5D-Y is a practical instrument to implement in primary care to routine monitoring. It is a promising instrument to compare the efficiency of different programs or treatment strategies, helping prioritize and invest at different levels.

Acknowledgments

Financial support for this study was provided by the Instituto de Salud Carlos III FEDER: Fondo Europeo de Desarrollo Regional (PI12/01296 and FI16/00071) and DIUE of Generalitat de Catalunya (2017 SGR 452). The funding agreements ensure the authors' independence in designing the study, interpreting the data, and writing and publishing the report. The authors would like to acknowledge Aurea Martin for her help in the English editing and proofreading process and in finalizing this manuscript. Author LR has now retired from the Agency for Health Quality & Assessment of Catalonia.

Authors' Contributions

KM contributed to the conception and design of the article, conceptualized and oversaw analyses, contributed to the interpretation of data, and wrote the article; LR conceived the study, participated in the design and coordination, and carried out the fieldwork; MM participated in the design and coordination of the study and carried out the fieldwork; OG contributed to the conception and design of the article, conceptualized and oversaw analyses, and contributed to the interpretation of data; AP contributed to the analysis and provided statistical support; JA contributed to the conception and design of the article, conceptualized and oversaw analyses, and contributed to the interpretation of data; JB participated in the design of the study and carried out the fieldwork; JP, RC, GC, XH, JMM, and DRA participated in the design of the study and carried out the fieldwork; URS conceived the study and participated in the design and coordination of the study; HR conceived the study and participated in the coordination of the study; VSS revised important intellectual content and the draft versions of the manuscript; MF oversaw all aspects, contributed to the conception and design of the article, contributed to the statistical analyses, carried out the interpretation of data, and contributed to the writing of the article. All the coauthors critically revised the manuscript and approved the final draft before submission, can attest to the validity and legitimacy of the data in the manuscript, and agree to be named as author of the manuscript.

Conflicts of Interest

None declared.

References

1. Kalyva E, Malakonaki E, Eiser C, Mamoulakis D. Health-related quality of life (HRQoL) of children with type 1 diabetes mellitus (T1DM): self and parental perceptions. Pediatr Diabetes 2011 Feb;12(1):34-40. [doi: 10.1111/j.1399-5448.2010.00653.x] [Medline: 20546163]
2. American Diabetes Association. 13. Children and Adolescents: Standards of Medical Care in Diabetes-2019. Diabetes Care 2019 Jan;42(Suppl 1):S148-S164. [doi: 10.2337/dc19-S013] [Medline: 30559239]
3. Mcdowell I. Factors affecting quality of life in lower limb amputees. In: Measuring Health: A Guide To Rating Scales And Questionnaires. Oxford, UK: Oxford University Press; 2019.
4. Guyatt GH, Feeny DH, Patrick DL. Measuring health-related quality of life. Ann Intern Med 1993 Apr 15;118(8):622-629. [doi: 10.7326/0003-4819-118-8-199304150-00009] [Medline: 8452328]
5. Debono M, Cachia E. The impact of diabetes on psychological well being and quality of life. The role of patient education. Psychol Health Med 2007 Oct;12(5):545-555. [doi: 10.1080/13548500701235740] [Medline: 17828675]
6. van Reenen M, Janssen B, Oppe M, Kreimeier S, Greiner W. EuroQol. 2014 Aug. EQ-5D-Y User Guide: Basic information on how to use the EQ-5D-Y instrument URL: https://euroqol.org/wp-content/uploads/2019/10/EQ-5D-Y-User-Guide.pdf
7. Aaronson N, Alonso J, Burnam A, Lohr KN, Patrick DL, Perrin E, et al. Assessing health status and quality-of-life instruments: attributes and review criteria. Qual Life Res 2002 May;11(3):193-205. [Medline: 12074258]
8. Eiser C, Morse R. A review of measures of quality of life for children with chronic illness. Arch Dis Child 2001 Mar;84(3):205-211 [FREE Full text] [doi: 10.1136/adc.84.3.205] [Medline: 11207164]
9. Gusi N, Perez-Sousa MA, Gozalo-Delgado M, Olivares PR. Validity and reliability of the spanish EQ-5D-Y proxy version. An Pediatr (Barc) 2014 Oct;81(4):212-219 [FREE Full text] [doi: 10.1016/j.anpedi.2013.11.028] [Medline: 24411558]
10. Raat H, Bonsel GJ, Hoogeveen WC, Essink-Bot M, Dutch HUI Group. Feasibility and reliability of a mailed questionnaire to obtain visual analogue scale valuations for health states defined by the Health Utilities Index Mark 3. Med Care 2004 Jan;42(1):13-18. [doi: 10.1097/01.mlr.0000102297.06535.e7] [Medline: 14713735]
11. Ravens-Sieberer U, Wille N, Badia X, Bonsel G, Burström K, Cavrini G, et al. Feasibility, reliability, and validity of the EQ-5D-Y: results from a multinational study. Qual Life Res 2010 Aug;19(6):887-897 [FREE Full text] [doi: 10.1007/s11136-010-9649-x] [Medline: 20401552]
12. Burström K, Svartengren M, Egmar A. Testing a Swedish child-friendly pilot version of the EQ-5D instrument--initial results. Eur J Public Health 2011 Apr;21(2):178-183. [doi: 10.1093/eurpub/ckq042] [Medline: 20430806]
13. Jelsma J. A comparison of the performance of the EQ-5D and the EQ-5D-Y health-related quality of life instruments in South African children. Int J Rehabil Res 2010 Jun;33(2):172-177. [doi: 10.1097/MRR.0b013e32832bb0f7] [Medline: 20351650]
14. Jelsma J, Ramma L. How do children at special schools and their parents perceive their HRQoL compared to children at open schools? Health Qual Life Outcomes 2010 Jul 21;8:72 [FREE Full text] [doi: 10.1186/1477-7525-8-72] [Medline: 20663151]
15. Scott D, Ferguson GD, Jelsma J. The use of the EQ-5D-Y health related quality of life outcome measure in children in the Western Cape, South Africa: psychometric properties, feasibility and usefulness - a longitudinal, analytical study. Health Qual Life Outcomes 2017 Jan 19;15(1):12 [FREE Full text] [doi: 10.1186/s12955-017-0590-3] [Medline: 28103872]
16. Robles N, Rajmil L, Rodriguez-Arjona D, Azuara M, Codina F, Raat H, et al. Development of the web-based Spanish and Catalan versions of the Euroqol 5D-Y (EQ-5D-Y) and comparison of results with the paper version. Health Qual Life Outcomes 2015 Jun 03;13:72 [FREE Full text] [doi: 10.1186/s12955-015-0271-z] [Medline: 26037720]
17. Eidt-Koch D, Mittendorf T, Greiner W. Cross-sectional validity of the EQ-5D-Y as a generic health outcome instrument in children and adolescents with cystic fibrosis in Germany. BMC Pediatr 2009 Aug 28;9:55 [FREE Full text] [doi: 10.1186/1471-2431-9-55] [Medline: 19715563]
18. Bergfors S, Åström M, Burström K, Egmar AC. Measuring health-related quality of life with the EQ-5D-Y instrument in children and adolescents with asthma. Acta Paediatr 2015 Mar;104(2):167-173. [doi: 10.1111/apa.12863] [Medline: 25393977]
19. Hsu CN, Lin HW, Pickard AS, Tain YL. EQ-5D-Y for the assessment of health-related quality of life among Taiwanese youth with mild-to-moderate chronic kidney disease. Int J Qual Health Care 2018 May 01;30(4):298-305. [doi: 10.1093/intqhc/mzy011] [Medline: 29447362]
20. Miranda Velasco MJ, Domínguez Martín E, Arroyo Díez FJ, Méndez Pérez P, González de Buitrago Amigo J. Health related quality of life in type 1 diabetes mellitus. An Pediatr (Barc) 2012 Nov;77(5):329-333 [FREE Full text] [doi: 10.1016/j.anpedi.2012.03.005] [Medline: 22542635]
21. Murillo M, Bel J, Pérez J, Corripio R, Carreras G, Herrero X, et al. Health-related quality of life (HRQOL) and its associated factors in children with Type 1 Diabetes Mellitus (T1DM). BMC Pediatr 2017 Jan 13;17(1):16 [FREE Full text] [doi: 10.1186/s12887-017-0788-x] [Medline: 28086765]

22. Murillo M, Bel J, Pérez J, Corripio R, Carreras G, Herrero X, et al. Impact of monitoring health-related quality of life in clinical practice in children with type 1 diabetes mellitus. Qual Life Res 2017 Dec;26(12):3267-3277. [doi: 10.1007/s11136-017-1682-6] [Medline: 28803331]
23. Kreimeier S, Greiner W. EQ-5D-Y as a Health-Related Quality of Life Instrument for Children and Adolescents: The Instrument's Characteristics, Development, Current Use, and Challenges of Developing Its Value Set. Value Health 2019 Jan;22(1):31-37 [FREE Full text] [doi: 10.1016/j.jval.2018.11.001] [Medline: 30661631]
24. Wilke CT, Pickard AS, Walton SM, Moock J, Kohlmann T, Lee TA. Statistical implications of utility weighted and equally weighted HRQL measures: an empirical study. Health Econ 2010 Jan;19(1):101-110. [doi: 10.1002/hec.1467] [Medline: 19248149]
25. Prieto L, Sacristán JA. What is the value of social values? The uselessness of assessing health-related quality of life through preference measures. BMC Med Res Methodol 2004 Apr 29;4:10 [FREE Full text] [doi: 10.1186/1471-2288-4-10] [Medline: 15117417]
26. Hinz A, Kohlmann T, Stöbel-Richter Y, Zenger M, Brähler E. The quality of life questionnaire EQ-5D-5L: psychometric properties and normative values for the general German population. Qual Life Res 2014 Mar;23(2):443-447. [doi: 10.1007/s11136-013-0498-2] [Medline: 23921597]
27. Ravens-Sieberer U, Erhart M, Rajmil L, Herdman M, Auquier P, Bruil J, European KIDSCREEN Group. Reliability, construct and criterion validity of the KIDSCREEN-10 score: a short measure for children and adolescents' well-being and health-related quality of life. Qual Life Res 2010 Dec;19(10):1487-1500 [FREE Full text] [doi: 10.1007/s11136-010-9706-5] [Medline: 20668950]
28. Ravens-Sieberer U, Auquier P, Erhart M, Gosch A, Rajmil L, Bruil J, European KIDSCREEN Group. The KIDSCREEN-27 quality of life measure for children and adolescents: psychometric results from a cross-cultural survey in 13 European countries. Qual Life Res 2007 Oct;16(8):1347-1356. [doi: 10.1007/s11136-007-9240-2] [Medline: 17668292]
29. Goodman R. The Strengths and Difficulties Questionnaire: a research note. J Child Psychol Psychiatry 1997 Jul;38(5):581-586. [Medline: 9255702]
30. Cladellas Y. Dep Psiquiatr i Med Leg Fac Med. 2007. El desenvolupament de la personalitat psicopàtica: pràctiques educatives parentals i altres variables familiars URL: https://ddd.uab.cat/record/38096 [accessed 2019-11-04]
31. Lima Rodríguez JS, Lima Serrano M, Jiménez Picón N, Domínguez Sánchez I. Reliability and construct validity of an instrument to asses the Self-perception of Family Health Status [in Spanish]. Rev Esp Salud Publica 2012 Oct;86(5):509-521 [FREE Full text] [doi: 10.4321/S1135-57272012000500005] [Medline: 23223763]
32. Smilkstein G, Ashworth C, Montano D. Validity and reliability of the family APGAR as a test of family function. J Fam Pract 1982 Aug;15(2):303-311. [Medline: 7097168]
33. Rosner B. Fundamentals of Biostatistics. Boston, MA: Cengage Learning; 2019.
34. Sullivan GM, Feinn R. Using Effect Size-or Why the P Value Is Not Enough. J Grad Med Educ 2012 Sep;4(3):279-282 [FREE Full text] [doi: 10.4300/JGME-D-12-00156.1] [Medline: 23997866]
35. Cohen J. Statistical power analysis for the behavioral sciences. Abingdon, United Kingdom: Routledge; 1988:1-567.
36. Davis RE, Morrissey M, Peters JR, Wittrup-Jensen K, Kennedy-Martin T, Currie CJ. Impact of hypoglycaemia on quality of life and productivity in type 1 and type 2 diabetes. Curr Med Res Opin 2005 Sep;21(9):1477-1483. [doi: 10.1185/030079905X61929] [Medline: 16197667]
37. Wagner VM, Müller-Godeffroy E, von Sengbusch S, Häger S, Thyen U. Age, metabolic control and type of insulin regime influences health-related quality of life in children and adolescents with type 1 diabetes mellitus. Eur J Pediatr 2005 Aug;164(8):491-496. [doi: 10.1007/s00431-005-1681-4] [Medline: 15875213]
38. Kristensen LJ, Birkebaek NH, Mose AH, Hohwü L, Thastum M. Symptoms of emotional, behavioral, and social difficulties in the danish population of children and adolescents with type 1 diabetes--results of a national survey. PLoS One 2014;9(5):e97543 [FREE Full text] [doi: 10.1371/journal.pone.0097543] [Medline: 24842772]
39. Delamater AM, Shaw KH, Applegate EB, Pratt IA, Eidson M, Lancelotta GX, et al. Risk for metabolic control problems in minority youth with diabetes. Diabetes Care 1999 May;22(5):700-705 [FREE Full text] [doi: 10.2337/diacare.22.5.700] [Medline: 10332669]
40. Silverstein J, Cheng P, Ruedy KJ, Kollman C, Beck RW, Klingensmith GJ, Pediatric Diabetes Consortium. Depressive Symptoms in Youth With Type 1 or Type 2 Diabetes: Results of the Pediatric Diabetes Consortium Screening Assessment of Depression in Diabetes Study. Diabetes Care 2015 Dec;38(12):2341-2343. [doi: 10.2337/dc15-0982] [Medline: 26459274]
41. Burnand B, Kernan WN, Feinstein AR. Indexes and boundaries for. J Clin Epidemiol 1990;43(12):1273-1284. [doi: 10.1016/0895-4356(90)90093-5] [Medline: 2254764]
42. Terwee CB, Bot SDM, de Boer MR, van der Windt DAWM, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 2007 Jan;60(1):34-42. [doi: 10.1016/j.jclinepi.2006.03.012] [Medline: 17161752]
43. Sprangers MA, Schwartz CE. Integrating response shift into health-related quality of life research: a theoretical model. Soc Sci Med 1999 Jun;48(11):1507-1515. [Medline: 10400253]
44. Lustman PJ, Freedland KE, Carney RM, Hong BA, Clouse RE. Similarity of depression in diabetic and psychiatric patients. Psychosom Med 1992;54(5):602-611. [doi: 10.1097/00006842-199209000-00007] [Medline: 1438662]

45. Wake M, Hesketh K, Cameron F. The Child Health Questionnaire in children with diabetes: cross-sectional survey of parent and adolescent-reported functional health status. Diabet Med 2000 Oct;17(10):700-707. [doi: 10.1046/j.1464-5491.2000.00360.x] [Medline: 11110502]
46. Zenlea IS, Mednick L, Rein J, Quinn M, Wolfsdorf J, Rhodes ET. Routine behavioral and mental health screening in young children with type 1 diabetes mellitus. Pediatr Diabetes 2014 Aug;15(5):384-388 [FREE Full text] [doi: 10.1111/pedi.12099] [Medline: 24274235]
47. Hoey H, Aanstoot HJ, Chiarelli F, Daneman D, Danne T, Dorchy H, et al. Good metabolic control is associated with better quality of life in 2,101 adolescents with type 1 diabetes. Diabetes Care 2001 Nov;24(11):1923-1928. [doi: 10.2337/diacare.24.11.1923] [Medline: 11679458]
48. Ismail H. Self-rated health and factors influencing responses among young Egyptian type 1 diabetes patients. BMC Public Health 2011 Apr 07;11:216 [FREE Full text] [doi: 10.1186/1471-2458-11-216] [Medline: 21473776]
49. Garratt AM, Schmidt L, Fitzpatrick R. Patient-assessed health outcome measures for diabetes: a structured review. Diabet Med 2002 Jan;19(1):1-11. [doi: 10.1046/j.1464-5491.2002.00650.x] [Medline: 11869297]
50. Ravens-Sieberer U, Herdman M, Devine J, Otto C, Bullinger M, Rose M, et al. The European KIDSCREEN approach to measure quality of life and well-being in children: development, current application, and future advances. Qual Life Res 2014 Apr;23(3):791-803 [FREE Full text] [doi: 10.1007/s11136-013-0428-3] [Medline: 23686556]
51. Desai AD, Zhou C, Stanford S, Haaland W, Varni JW, Mangione-Smith RM. Validity and responsiveness of the pediatric quality of life inventory (PedsQL) 4.0 generic core scales in the pediatric inpatient setting. JAMA Pediatr 2014 Dec;168(12):1114-1121. [doi: 10.1001/jamapediatrics.2014.1600] [Medline: 25347549]
52. Landgraf JM, van Grieken A, Raat H. Giving voice to the child perspective: psychometrics and relative precision findings for the Child Health Questionnaire self-report short form (CHQ-CF45). Qual Life Res 2018 Aug;27(8):2165-2176. [doi: 10.1007/s11136-018-1873-9] [Medline: 29777459]

Abbreviations

APGAR: Adaptation, Partnership, Growth, Affection, and Resolve
DUA: doing usual activities
EQ-5D-Y: Youth EuroQoL version
EQ-VAS: EuroQol-Visual Analog Scale
EQ: EuroQol
ES: effect size
HbA_{1c}: glycated hemoglobin
HP/D: having pain/discomfort
HRQoL: health-related quality of life
ICC: intraclass correlation coefficient
LAM: looking after myself
SDQ: Strengths and Difficulties Questionnaire
T1DM: type I diabetes mellitus
W/S/U: feeling worried/sad/unhappy

Edited by G Eysenbach; submitted 07.06.19; peer-reviewed by G Hu, T Oser; comments to author 31.08.19; revised version received 13.09.19; accepted 24.09.19; published 12.11.19.

Please cite as:
Mayoral K, Rajmil L, Murillo M, Garin O, Pont A, Alonso J, Bel J, Perez J, Corripio R, Carreras G, Herrero J, Mengibar JM, Rodriguez-Arjona D, Ravens-Sieberer U, Raat H, Serra-Sutton V, Ferrer M
Measurement Properties of the Online EuroQol-5D-Youth Instrument in Children and Adolescents With Type 1 Diabetes Mellitus: Questionnaire Study
J Med Internet Res 2019;21(11):e14947
URL: http://www.jmir.org/2019/11/e14947/
doi:10.2196/14947
PMID:31714252

Original Paper

A Scale to Assess the Methodological Quality of Studies Assessing Usability of Electronic Health Products and Services: Delphi Study Followed by Validity and Reliability Testing

Anabela G Silva[1], DPhil; Patrícia Simões[1], BSc; Rita Santos[2], DPhil; Alexandra Queirós[1], DPhil; Nelson P Rocha[3], DPhil; Mário Rodrigues[2], DPhil

[1]School of Health Sciences, University of Aveiro, Aveiro, Portugal

[2]Higher School of Technology and Management of Águeda, Aveiro, Portugal

[3]Department of Medical Sciences, University of Aveiro, Aveiro, Portugal

Corresponding Author:
Anabela G Silva, DPhil
School of Health Sciences
University of Aveiro
Agras do Crasto
Campus Universitário de Santiago
Aveiro,
Portugal
Phone: 351 234247119 ext 27120
Email: asilva@ua.pt

Abstract

Background: The usability of electronic health (eHealth) and mobile health apps is of paramount importance as it impacts the quality of care. Methodological quality assessment is a common practice in the field of health for different designs and types of studies. However, we were unable to find a scale to assess the methodological quality of studies on the usability of eHealth products or services.

Objective: This study aimed to develop a scale to assess the methodological quality of studies assessing usability of mobile apps and to perform a preliminary analysis of of the scale's feasibility, reliability, and construct validity on studies assessing usability of mobile apps, measuring aspects of physical activity.

Methods: A 3-round Delphi panel was used to generate a pool of items considered important when assessing the quality of studies on the usability of mobile apps. These items were used to write the scale and the guide to assist its use. The scale was then used to assess the quality of studies on usability of mobile apps for physical activity, and it assessed in terms of feasibility, interrater reliability, and construct validity.

Results: A total of 25 experts participated in the Delphi panel, and a 15-item scale was developed. This scale was shown to be feasible (time of application mean 13.10 [SD 2.59] min), reliable (intraclass correlation coefficient=0.81; 95% CI 0.55-0.93), and able to discriminate between low- and high-quality studies (high quality: mean 9.22 [SD 0.36]; low quality: mean 6.86 [SD 0.80]; *P*=.01).

Conclusions: The scale that was developed can be used both to assess the methodological quality of usability studies and to inform its planning.

(J Med Internet Res 2019;21(11):e14829) doi:10.2196/14829

KEYWORDS

quality of health care; eHealth; mHealth; efficiency

Introduction

Background

Methodological quality can be defined as "the extent to which study authors conducted their research to the highest possible standards" [1]. It should be considered both when interpreting individual study findings and when conducting systematic reviews and aggregating findings from different studies and making recommendations [1,2]. However, the critical assessment of the quality of studies is a complex process that must consider several different aspects of the study, which may vary depending on the type of study and on the subject of research [1,3,4]. Therefore, this process is usually performed with the aid of critical appraisal tools previously developed for that specific purpose. This is common practice in the field of health, where a number of critical appraisal tools exist to assist the assessment of the methodological quality of studies [1,3-5]. There are different tools depending, for example, on whether studies are randomized clinical trials aiming to assess the effectiveness of interventions [5] or assess the validity and/or reliability of measurement instruments [6] or are diagnostic accuracy studies [3]. However, we were unable to find any critical tool to guide the assessment of methodological quality of usability studies, neither for electronic health (eHealth) applications nor for general applications.

According to the International Standards Organization 9241-11, usability refers to the "extent to which a system, product or service can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use" [7]. Therefore, usability evaluation is an important part of the process of development of any system, product, or service [8] and can be formative or summative, that is, its main focus may be to detect and solve problems or to meet the metrics associated with the system, product, or service task and goals [9]. The complex nature of usability often requires the use of combined approaches for its assessment [8], involving, for example, the triangulation of methods, the use of both experts and end users, and different settings (eg, laboratory or real context). Furthermore, the type of instruments and procedures that are more adequate depend on several factors, such as the aim of the usability assessment (formative or summative) and on the development phase of the system, product, or service [10]. A methodologically sound assessment of usability is crucial to minimize the probability of errors and undesirable consequences and to increase the probability of use by a large proportion of the target end users [7]. In the field of health, usability contributes to enhance patient safety and quality of care, and recommendations aiming to enhance these by means of improving the usability have been published [11] as well as protocols to measure and validate user performance before deployment [11]. However, poor assessment of usability is common practice and impacts the quality of the eHealth apps [11]. Therefore, having a reference guide that could be used both to inform the design of usability studies and to assess the methodological quality of published studies is of paramount importance and constitutes a step forward in the field of usability.

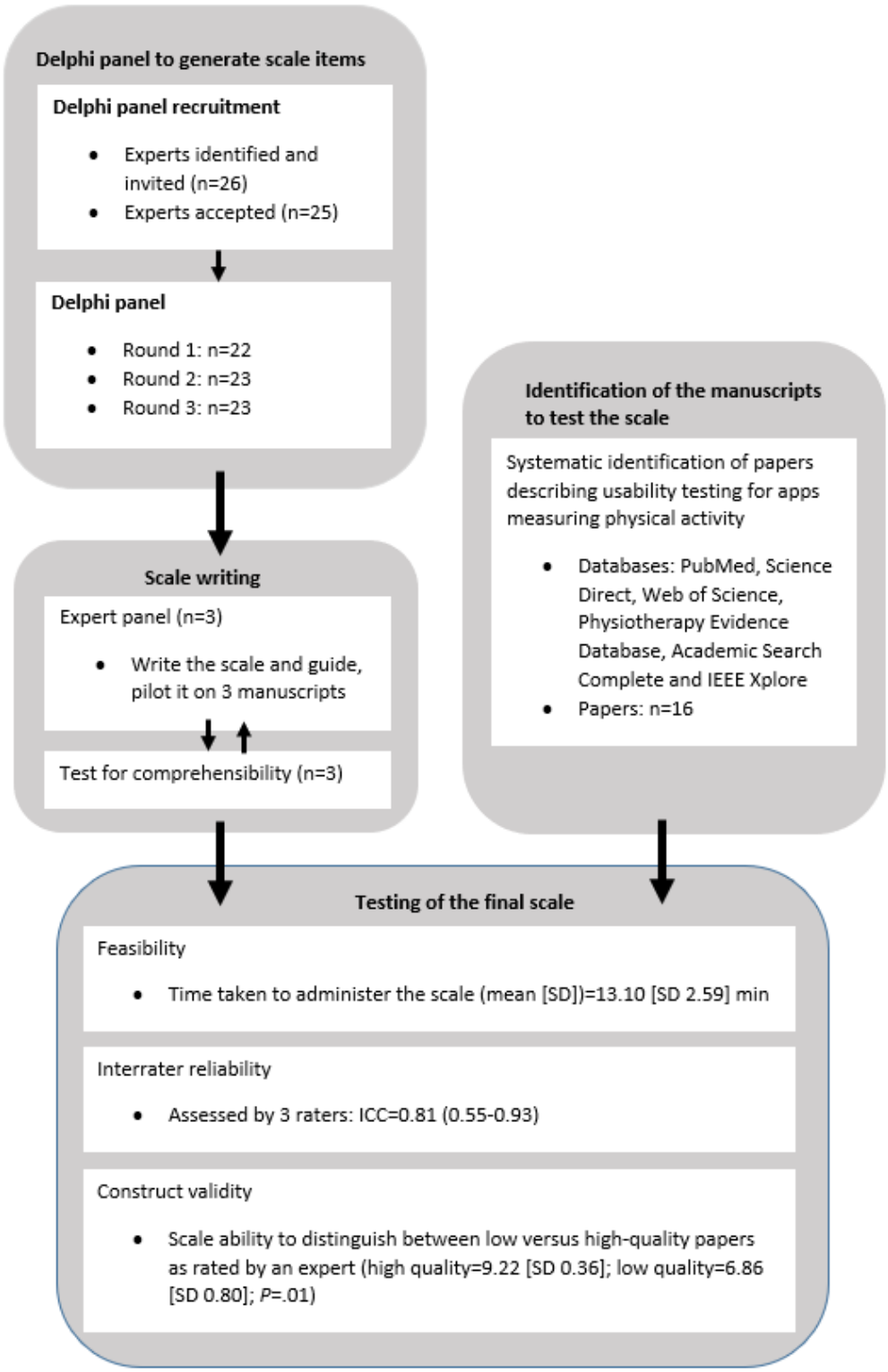
Objectives

The aims of this study were to develop a scale to assess the methodological quality of studies assessing usability of mobile apps and to perform a preliminary analysis of its feasibility, reliability, and construct validity on studies assessing usability of mobile apps, measuring aspects of physical activity.

Methods

This study comprised 3 phases: (1) a 3-round Delphi panel to generate a pool of items considered important when assessing the quality of studies on usability; (2) a panel of experts to write the scale and the guide to assist in the use of the scale when assessing the quality of studies on usability; (3) testing of the developed scale, including the assessment of feasibility, interrater reliability, and construct validity. Figure 1 shows the flow of procedures for this study.

Figure 1. Flowchart of study procedures. ICC: intraclass correlation coefficient.



Delphi Panel

The Delphi method was used because it is recommended to determine the consensus for a predefined problem when there is little information, and one must rely on the opinion of experts. It is a structured multistage process to collect information from an expert panel about a certain topic to reach consensus based on structured group communication [12].

Expert Selection

To take part in the study, experts had to meet the following criteria: (1) have experience conducting studies on usability and (2) have previous published work on assessment of usability. In addition, we aimed to recruit participants with diverse academic backgrounds (Technology, Health, and Design) so that different points of view could be gathered. Participants who complied with these criteria were identified by team members. A sample size of at least 20 experts has been suggested as

appropriate [12,13]. To account for potential dropouts through the rounds, a total of 26 experts were invited to enter the study by an individual email or a phone call, and 25 of them accepted. Experts' anonymity was maintained throughout the study.

Development of Consensus

This Delphi study was organized in 3 rounds. In the first round, participants were sent an email explaining the study with a link to a questionnaire developed using Google Forms and were asked to identify, by order of relevance, a minimum of 4 items that they thought were important to consider when assessing the quality of studies on usability. They were also asked to provide a justification for their choice.

The results of the first round were collated and then grouped into categories and subcategories with the same meaning so that at the end of this process, all items identified by all experts were allocated to a subcategory. This process was performed independently by 3 researchers (AGS, PS, and AR), who then met to compare their coding, and a consensus was reached. Each subcategory gave origin to a statement about an aspect that should be checked when assessing usability studies' methodological quality. The list of statements was sent back to experts in the second round. In this round, experts were asked to rate the relevance of each statement using a 9-item Likert scale (1—item not important to 9—item very important). Participants were also asked to give their opinion on the formulation of the items. Consensus on the inclusion of 1 item in the scale was considered when 70% or more participants scored the item as 7 to 9 and less than 15% of participants scored it as 1 to 3. Consensus on the exclusion of 1 item was considered when 70% or more participants scored the item as 1 to 3 and less than 15% of participants scored the item as 7 to 9 [14,15]. The changes recommended by experts on the writing of each item, and which were considered relevant, were included in the third round.

In the third round, each expert has been presented with his/her previous score of each item and the ratings of the remaining experts summarized as absolute frequencies and presented in a graphic format. Experts were then asked whether they would like to reconsider their previous rating. The final list included all items that were classified with 7 to 9 regarding the degree of importance by at least 70% of the participants [14].

For each round, a minimum response rate of 70% was required to consider the round valid [16]. Experts had between 2 and 3 weeks to respond to each round, and reminders were sent at the end of the first and second weeks to those that had yet to reply.

Panel of Experts for Scale Writing

A total of 3 researchers (AGS, ARS, and PS) met to agree on the final writing of the scale to assess the quality of studies evaluating usability and how they should be ordered and prepared a first draft of the guide/manual of use to assist on using the scale. It was decided that an item should be scored as 0 if it was not assessed or not described in the study being appraised and as 1 if the item was assessed and that adding up the individual item score would result in a final global score. This first draft was piloted independently by the researchers on 3 manuscripts. In the second meeting, the panel revised the first draft of the scale based on their experience of using it. As it was decided that 2 items of the scale could be considered *not applicable*, we determined that the final score should be presented as percentage (ie, [number of items scored 1/total numbers of items applicable] × 100). This version of the scale was then sent to 3 experts external to this panel for comprehensibility assessment. These experts made only minor suggestions that were considered for inclusion, and the final version of the scale was named as Critical Assessment of Usability Studies Scale (CAUSS).

Feasibility, Interrater Reliability, and Construct Validity of the Final Scale

Feasibility and Interrater Reliability

To evaluate the feasibility and interrater reliability of the CAUSS, a systematic search for studies assessing usability of mobile apps measuring aspects of physical activity was performed on PubMed, Science Direct, Web of Science, Physiotherapy Evidence Database, Academic Search Complete, and IEEE Xplore using a combination of the following expressions: physical activity, mobile applications, and usability. All databases were searched since January 1, 2000, and the search was performed on October 29 and 30, 2017. We chose to use studies on the usability of mobile apps measuring aspects of physical activity to assess reliability as this study was conducted within the context of a research project on mobile (ie, eHealth or mobile health [mHealth]) apps to promote physical activity. To be included in the reliability part of this study, manuscripts had to (1) be full text; (2) specify the assessment of usability as one of its aims; and (3) describe the assessment of usability of an eHealth or mHealth app aiming primarily at measuring physical activity at any stage of development. A total of 16 studies met the inclusion criteria [17-32]. These studies were assessed independently by 3 authors (AGS, ARS, and PS) using the CAUSS and the respective guide. An intraclass correlation coefficient (ICC; 2-way random; absolute agreement) was used to compare the total score among raters, and an ICC of at least 0.7 was considered acceptable [33]. In addition, a repeated measures analysis of variance was also used to explore for significant differences between the scores of the 3 raters.

Feasibility was evaluated by assessing the time taken to assess the 16 studies using the CAUSS.

Construct Validity of the Scale

As there was no gold standard against which to compare the results of our scale, construct validity was assessed using a method adapted from Jadad et al [34] and Yates et al [35]. Articles were allocated to a group by a rater with extensive knowledge on both usability and methodological quality (NPR). This rater categorized each one of the 16 articles as low or high quality. Construct validity of the scale was assessed by testing whether it was able to discriminate between these categories. The consensus ratings of the 3 judges (AGS, ARS, and PS) who assessed each manuscript using the CAUSS (as detailed in the reliability section) were used for this analysis. A Student *t* test (data followed a normal distribution) was used to compare the scores of the articles classified as low and high quality.

Results

Delphi Panel

Of the 25 experts that entered the study, 11 (44%) were females, 21 (84%) held a Doctor of Philosophy degree, and their areas of academic background were diverse (Sciences and Technology of Communication, Engineering and Mathematics, Health, and Design; Table 1).

The first round was completed between April and June 2018, and a total of 22 out of 25 experts (88%) answered the questionnaire. In this round, the panel of experts generated a total of 121 statements where each person generated 5 statements on average (SD 0.87). The total statements were grouped in 22 main topics (Table 2), of which 6 were excluded by 3 members of the research team (Textbox 1) because they were not specific of usability studies and/or because they were out of scope (ie, not related to usability). The remaining topics were transformed into 15 questions and sent back to experts in round 2 (Table 2). Round 2 was completed by experts between November 2018 and January 2019. A total of 23 experts (92%) answered the questionnaire. Of the 15 questions identified, 13 reached consensus for inclusion.

The third and final round was completed between January and February 2019 by 23 out of 25 experts (92%). In this round, 14 of the 15 statements reached consensus and were included in the scale. However, the statement that did not reach consensus was also included because most of the experts (18/23, 78%) classified it with 6 or more out of a maximum score of 9. Table 3 shows the score of the 15 questions after rounds 2 and 3. These final statements were then used by the panel of 3 experts to write the final scale and its guide/manual of use (Multimedia Appendix 1).

Table 1. Characterization of experts participating in the Delphi panel (n=25).

Characteristics	Values
Gender, n (%)	
Male	14 (56)
Female	11 (44)
Age (years), median (IQR[a])	42 (14)
Education, n (%)	
Masters	4 (16)
Doctoral	21 (84)
Academic background, n (%)	
Sciences and Technology of Communication	7 (28)
Engineering and Mathematics	11 (44)
Health	5 (20)
Design	2 (8)
Current professional occupation, n (%)	
University lecturer	16 (64)
Researcher	8 (32)
Designer	1 (4)
Experience in usability assessment (years), median (IQR)	10 (10)

[a]IQR: interquartile range.

Table 2. Subcategories generated after round 1 and included in round 2.

Subcategories	Questions sent back to experts in round 2
Valid measurement instruments	Did the study use valid measurement instruments of usability (ie, there is evidence that the instruments used assess usability)?
Reliable measurement instruments	Did the study use reliable measurement instruments of usability (ie, there is evidence that the instruments used have similar results in repeated measurements in similar circumstances)?
Procedures adequate to the study's objectives	Was there coherence between the procedures used to assess usability (eg, instruments and context) and study aims?
Procedures adequate to the development stage of the product	Did the study use procedures of assessment for usability that were adequate to the development stage of the product/service?
Procedures adequate to the participants' characteristics	Did the study use procedures of assessment for usability adequate to study participants' characteristics (eg, children and elderly require different instruments)?
Triangulation	Did the study employ triangulation of methods for the assessment of usability?
Combination of users' and experts' evaluation	Was usability assessed using both potential users and experts?
Experience of the investigator that conducted the usability evaluation	Was the investigator that conducted usability assessments adequately trained?
Investigator conducting usability assessment external to the development of the product/service	Was the investigator that conducted usability assessments external to the process of product/service development?
Assessment in real context or close to real context	Was the usability assessment conducted in the real context or close to the real context where product/service is going to be used?
Number of participants (potential users and/or experts)	Was the number of participants used to assess usability adequate (whether potential users or experts)?
Representativeness of participants (potential users and/or experts)	Were participants who assessed the product/service usability representative of the experts' population and/or of the potential users' population?
Representativeness of the tasks to perform on the usability evaluation	Were the tasks that serve as the base for the usability assessment representative of the functionalities of the product/service?
Continuous and prolonged use of the product	Was the usability assessment based on continuous and prolonged use of the product/service over time?
Analysis of the results	Was the type of analysis adequate to the study's aims and variables assessed?

Textbox 1. Subcategories generated after round 1 and not included in round 2.

1. Compliance with ethical principles
2. Pilot study before the main study
3. Definition of a protocol before study beginning
4. Description of study objectives, tasks, methods, measurement instruments, measures, context, and mobile app
5. The study is possible to replicate and/or reproduce
6. Others (negative impacts of usability, sample motivation, and development cycle)

Table 3. Results from rounds 2 and 3 of the Delphi panel.

Questions	Second round score, n (%)			Third round score, n (%)			Consensus
	1-3	4-6	7-9	1-3	4-6	7-9	
Did the study use valid measurement instruments of usability (ie, there is evidence that the instruments used assess usability)?	0 (0)	0 (0)	23 (100)	0 (0)	0 (0)	23 (100)	Yes
Did the study use reliable measurement instruments of usability (ie, there is evidence that the instruments used have similar results in repeated measurements in similar circumstances)?	0 (0)	3 (13)	20 (87)	0 (0)	0 (0)	23 (100)	Yes
Was there coherence between the procedures used to assess usability (eg, instruments, context) and study aims?	0 (0)	4 (17)	19 (83)	0 (0)	2 (9)	21 (91)	Yes
Did the study use procedures of assessment for usability that were adequate to the development stage of the product/service?	0 (0)	5 (22)	18 (78)	0 (0)	4 (17)	19 (83)	Yes
Did the study use procedures of assessment for usability adequate to study participants' characteristics (eg, children and elderly require different instruments)?	0 (0)	2 (9)	21 (91)	0 (0)	1 (4)	22 (96)	Yes
Did the study employ triangulation of methods for the assessment of usability?	0 (0)	5 (22)	18 (78)	0 (0)	6 (26)	17 (74)	Yes
Was usability assessed using both potential users and experts?	0 (0)	5 (22)	18 (78)	0 (0)	4 (17)	19 (83)	Yes
Were participants who assessed the product/service usability representative of the experts' population and/or of the potential users' population?	0 (0)	3 (13)	20 (87)	0 (0)	2 (9)	21 (91)	Yes
Was the investigator that conducted usability assessments adequately trained?	1 (4)	4 (17)	18 (78)	1 (4)	1 (4)	21 (91)	Yes
Was the investigator that conducted usability assessments external to the process of product/service development?	1 (4)	10 (44)	12 (52)	1 (4)	10 (43)	12 (52)	No[a]
Was the usability assessment conducted in the real context or close to the real context where product/service is going to be used?	0 (0)	5 (22)	18 (78)	0 (0)	2 (9)	21 (91)	Yes
Was the number of participants used to assess usability adequate (whether potential users or experts)?	0 (0)	2 (9)	21 (91)	0 (0)	0 (0)	23 (100)	Yes
Were the tasks that serve as the base for the usability assessment representative of the functionalities of the product/service?	0 (0)	0 (0)	23 (100)	0 (0)	0 (0)	23 (100)	Yes
Was the usability assessment based on continuous and prolonged use of the product/service over time?	0 (0)	9 (39)	14 (61)	0 (0)	6 (26)	17 (74)	Yes
Was the type of analysis adequate to the study's aims and variables assessed?	0 (0)	1 (4)	22 (96)	0 (0)	0 (0)	23 (100)	Yes

[a]This item was included because most of the experts (n=18, 78%) classified it with 6 or more out of a maximum score of 9.

Feasibility, Interrater Reliability, and Construct Validity of the Final Scale

Feasibility

The time taken (in minutes) to assess the articles using the scale varied between 10 and 18 min (mean 13.10 [SD 2.59] min).

Interrater Reliability

The 3 judges assessing the interrater reliability achieved an ICC of 0.81 (0.55-0.93) for the total scoring. Mean (SD) for the 3 raters was 8.63 (1.41), 8.60 (2.00), and 8.44 (1.50), and no significant difference was found between them ($F_{2,14}$=0.29; P=.75). Multimedia Appendix 2 presents the raters' score for each of the 15 items of the scale.

Construct Validity

The rater classified 9 articles as high quality and 7 articles as low quality. Mean (SD) of the scale's total score for the 2 groups of articles using the consensus score for each paper was significantly different: 9.22 (0.36) for the high-quality group and 6.86 (0.80) for the low-quality group (P=.01).

Discussion

This study presents a scale to assess the methodological quality of studies assessing usability, which was developed through a modified Delphi panel. Results of a pilot test of the scale on papers assessing usability of eHealth apps that measure physical activity suggest that the scale is feasible, valid, and reliable.

Validity

Content validity of the scale is supported by the consensus generated among a group of experts with diverse backgrounds and areas of expertise allowing us to capture a broad perspective on usability [32]. In addition, many of the methodological aspects of usability studies covered in the 15 items of the scale have been previously reported as relevant, such as validity and

reliability of the instruments used to assess usability, adequate sample size [9,10], combined use of different methods of usability assessment, and adequacy of study procedures to the development stage of the product/service [36].

Further evidence on the scale validity comes from the fact that general items such as reliability and validity of instruments used, adequate sample size, competence of the assessor, appropriateness of analysis methods, or representativeness of participants are also reported in other scales [3,4,6,34] and from the scale's ability to distinguish between low- and high-quality trials (construct validity).

Reliability

The interrater reliability was acceptable, but the lower limit of the confidence interval is below the cut off for acceptable reliability. The raters involved in reliability testing had diverse backgrounds (health and engineering) and different degrees of experience rating the methodological quality of studies, which may have had an impact on the reliability results. Items 6, 7, 9, and 13 were the items of the scale where disagreement was more frequent. The lack of detail of the Methods section of the papers assessed and the different degrees of expertise of the raters on quantitative and qualitative data analysis may help explain why disagreement was more marked for these items. Furthermore, and for item 9 (participants representative of the experts' population and/or of the potential users' population), a few questions arose during the discussion to reach consensus among the 3 raters, particularly regarding the minimal set of characteristics that study authors need to provide to allow the reader/assessor to be able to judge on whether study participants were representative. For example, the users' age, sex, and previous experience using mobile phones and apps are important aspects to consider when judging the representativeness of the sample [36]. Similarly, a low ratio between the initial number of participants invited and those that entered the study as well as a less optimal recruitment process can lead to a more homogeneous sample with specific characteristics, which is less likely to be representative of the wider population [37]. For experts, area of expertise, years of practice, and previous experience using similar applications are examples of relevant characteristics to consider. However, a clear and complete description of participants, either experts or potential users, was generally not given in the studies assessed. These aspects may have contributed to the lower agreement on the referred items.

In contrast, items 8 (use of both potential users and experts), 11 (was the investigator that conducted usability assessments external), and 15 (continuous and prolonged use of the product/service) were consensual for all studies. Interestingly, all studies (except one) received the same rating for items 8 and 9 (insufficient information provided by study authors/no) and 15 (yes). The apparent higher objectivity of these items, the absence of information on who was the person conducting usability assessments, and the clear description of the period during which the application was used may explain the higher agreement between raters for these items.

Identified Shortcomings of the Papers Assessed

There were several aspects that were consistently not considered or for which insufficient detail was provided by authors of the papers assessed using our scale. Reliability and validity of the instruments used were never reported in the papers assessed. When this item was rated as "yes," meaning that studies employed reliable and/or valid instruments, it was because the instruments used were known to be valid and reliable. For data collected using qualitative methodologies, there was insufficient detail on how the analysis was conducted and how many researchers were involved. Using valid instruments (ie, instruments that measure what they are expected to measure) and instruments that are reliable (ie, instruments that give consistent ratings in the same conditions) are fundamental so that one can trust on the results of the assessment [38]. Information regarding who was the person performing usability assessments and previous experience and/or training to perform usability assessment was seldom given. However, previous experience or adequate training is fundamental, particularly for qualitative assessments of usability, and having an interest on the service/product being tested may bias the results. This has been shown in the field of health whether beliefs and expectations have been found to have an impact on the study results [39]. The lack of clarity of reports on usability assessment has already been pointed by other authors [40]. Future studies assessing usability of products/services should clearly report on these details. Poor reporting may reflect the lack of planning and poor methodological quality.

Limitations

The underlying assumption of calculating the total score of CAUSS by simply adding individual items is that all items are equally important to the final score. This is the simplest and most commonly used solution, but it does not account for the varying relevance of individual items to the construct being measured [41]. In contrast, adding up items makes the scale easier to score and, potentially, more appealing for use. Nevertheless, it could be argued that the 15 items of the CAUSS are not all equality relevant in terms of the methodological quality of usability studies. The impact of using different methods to calculate the final score could be explored in future studies aiming at further refinement of the scale.

Reliability was assessed only by researchers involved in the development of the scale, which may have inflated the reliability results. In addition, we assessed interrater reliability only and did not test for test-retest reliability, which assesses the consistency of ratings for the same rater. Nevertheless, test-retest reliability is usually higher than interrater reliability, as interrater reliability refers to intersubject variability which is usually higher than intrasubject variability. The limited number of experts used to assess validity and the absence of other scales assessing the methodological quality of usability studies limit our ability to compare results. The future use of the developed scale to assess the methodological quality of other products/service will provide data on the reliability and validity of the scale. We encourage researchers to use the scale and to provide feedback.

In summary, the CAUSS scale, developed to assess the methodological quality of studies assessing usability, seems to be feasible to use and to have construct validity and interrater reliability. Further reliability, including test-retest reliability, and validity testing should be performed for different products and services, and the impact of using different methods to calculate the final score should also be explored.

Acknowledgments

This study was partially funded by SmartWalk—Smart Cities for Active Seniors (CENTRO-01-0145-FEDER-024293).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Guide for the methodological quality assessment of studies that evaluate usability.
[PDF File (Adobe PDF File), 67 KB - jmir_v21i11e14829_app1.pdf]

Multimedia Appendix 2
Methodological assessment of studies for the reliability and validity analysis of the scale. NA: not applicable; consensus score: final score after consensus; %: percentage out of maximum possible score.
[PDF File (Adobe PDF File), 79 KB - jmir_v21i11e14829_app2.pdf]

References

1. Higgins JP, Green S. The Cochrane Collaboration.: The Cochrane Collaboration; 2011. Cochrane Handbook for Systematic Reviews of Interventions. URL: http://handbook-5-1.cochrane.org/ [accessed 2019-03-30]
2. Moja LP, Telaro E, Amico RD, Moschetti I, Coe L, Liberati A. Assessment of methodological quality of primary studies by systematic reviews: results of the metaquality cross sectional study. Br Med J 2005 May 7;330(7499):1053 [FREE Full text] [doi: 10.1136/bmj.38414.515938.8F] [Medline: 15817526]
3. Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Med Res Methodol 2003 Nov 10;3:25 [FREE Full text] [doi: 10.1186/1471-2288-3-25] [Medline: 14606960]
4. Wells G, Shea B, O'Connell D, Peterson J, Welch V, Losos M, et al. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. Eur J Epidemiol 2011;25:603-605 [FREE Full text]
5. Olivo SA, Macedo LG, Gadotti IC, Fuentes J, Stanton T, Magee DJ. Scales to assess the quality of randomized controlled trials: a systematic review. Phys Ther 2008 Feb;88(2):156-175. [doi: 10.2522/ptj.20070147] [Medline: 18073267]
6. Brink Y, Louw Q. Clinical instruments: reliability and validity critical appraisal. J Eval Clin Pract 2012 Dec;18(6):1126-1132. [doi: 10.1111/j.1365-2753.2011.01707.x] [Medline: 21689217]
7. International Standards Organization. 2019. Ergonomics of human-system interaction — Part 210: Human-centered design for interactive systems. URL: https://www.iso.org/obp/ui/#iso:std:iso:9241:-210:ed-2:v1:en [accessed 2019-07-20]
8. Martins AI, Queirós A, Silva AG, Rocha NP. Usability evaluation methods: a systematic review. In: Saeed S, Bajwa IM, Mahmood Z, editors. Human Factors in Software Development and Design. Pennsylvania: IGI Global; 2014.
9. Lewis JR. Usability: lessons learned … and yet to be learned. Int J Hum Comput Interact 2014 Jun 18;30(9):663-684. [doi: 10.1080/10447318.2014.930311]
10. Klaassen B, van Beijnum B, Hermens H. Usability in telemedicine systems-A literature survey. Int J Med Inform 2016 Sep;93:57-69. [doi: 10.1016/j.ijmedinf.2016.06.004] [Medline: 27435948]
11. Middleton B, Bloomrosen M, Dente MA, Hashmat B, Koppel R, Overhage JM, American Medical Informatics Association. Enhancing patient safety and quality of care by improving the usability of electronic health record systems: recommendations from AMIA. J Am Med Inform Assoc 2013 Jun;20(e1):e2-e8 [FREE Full text] [doi: 10.1136/amiajnl-2012-001458] [Medline: 23355463]
12. Okoli C, Pawlowski SD. The Delphi method as a research tool: an example, design considerations and applications. Inf Manag 2004 Dec;42(1):15-29. [doi: 10.1016/j.im.2003.11.002]
13. Skulmoski GJ, Hartman FT, Krahn J. The Delphi method for graduate research. J Inf Technol Educ 2007;6:1-21. [doi: 10.28945/199]
14. Guyatt GH, Oxman AD, Kunz R, Atkins D, Brozek J, Vist G, et al. GRADE guidelines: 2. Framing the question and deciding on important outcomes. J Clin Epidemiol 2011 Apr;64(4):395-400. [doi: 10.1016/j.jclinepi.2010.09.012] [Medline: 21194891]
15. Williamson P, Altman D, Blazeby J, Clarke M, Devane D, Gargon E, et al. Developing core outcome sets for clinical trials: issues to consider. Trials 2012 Aug 6;13:132 [FREE Full text] [doi: 10.1186/1745-6215-13-132] [Medline: 22867278]

16. Akins RB, Tolson H, Cole BR. Stability of response characteristics of a Delphi panel: application of bootstrap data expansion. BMC Med Res Methodol 2005 Dec 1;5:37 [FREE Full text] [doi: 10.1186/1471-2288-5-37] [Medline: 16321161]
17. Vorrink SN, Kort HS, Troosters T, Lammers JW. A mobile phone app to stimulate daily physical activity in patients with chronic obstructive pulmonary disease: development, feasibility, and pilot studies. JMIR Mhealth Uhealth 2016 Jan 26;4(1):e11 [FREE Full text] [doi: 10.2196/mhealth.4741] [Medline: 26813682]
18. Al Ayubi SU, Parmanto B, Branch R, Ding D. A persuasive and social mHealth application for physical activity: a usability and feasibility study. JMIR Mhealth Uhealth 2014 May 22;2(2):e25 [FREE Full text] [doi: 10.2196/mhealth.2902] [Medline: 25099928]
19. Soto-Perez-De-Celis E, Kim H, Rojo-Castillo MP, Sun C, Chavarri-Guerra Y, Navarrete-Reyes AP, et al. A pilot study of an accelerometer-equipped smartphone to monitor older adults with cancer receiving chemotherapy in Mexico. J Geriatr Oncol 2018 Mar;9(2):145-151. [doi: 10.1016/j.jgo.2017.09.008] [Medline: 29017891]
20. Sannino G, Forastiere M, de Pietro G. A wellness mobile application for smart health: pilot study design and results. Sensors (Basel) 2017 Mar 17;17(3):pii: E611 [FREE Full text] [doi: 10.3390/s17030611] [Medline: 28304332]
21. Hartzler AL, Venkatakrishnan A, Mohan S, Silva M, Lozano P, Ralston JD, et al. Acceptability of a Team-based Mobile Health (mHealth) Application for Lifestyle Self-management in Individuals With Chronic Illnesses. In: Proceedings of the 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2016 Presented at: EMBC'16; 2016; Orlando, Florida USA p. 3277-3281. [doi: 10.1109/EMBC.2016.7591428]
22. Mercer K, Giangregorio L, Schneider E, Chilana P, Li M, Grindrod K. Acceptance of commercially available wearable activity trackers among adults aged over 50 and with chronic illness: a mixed-methods evaluation. JMIR Mhealth Uhealth 2016 Jan 27;4(1):e7 [FREE Full text] [doi: 10.2196/mhealth.4225] [Medline: 26818775]
23. Oliver E, Baños RM, Cebolla A, Lurbe E, Alvarez-Pitti J, Botella C. An electronic system (PDA) to record dietary and physical activity in obese adolescents; data about efficiency and feasibility. Nutr Hosp 2013 Nov 1;28(6):1860-1866 [FREE Full text] [doi: 10.3305/nutr hosp.v28in06.6784] [Medline: 24506361]
24. Rospo G, Valsecchi V, Bonomi AG, Thomassen IW, van Dantzig S, La Torre A, et al. Cardiorespiratory improvements achieved by American college of sports medicine's exercise prescription implemented on a mobile app. JMIR Mhealth Uhealth 2016 Jun 23;4(2):e77 [FREE Full text] [doi: 10.2196/mhealth.5518] [Medline: 27339153]
25. Dunton GF, Dzubur E, Kawabata K, Yanez B, Bo B, Intille S. Development of a smartphone application to measure physical activity using sensor-assisted self-report. Front Public Heal 2014;2:12 [FREE Full text] [doi: 10.3389/fpubh.2014.00012] [Medline: 24616888]
26. Choo S, Kim JY, Jung SY, Kim S, Kim JE, Han JS, et al. Development of a weight loss mobile app linked with an accelerometer for use in the clinic: usability, acceptability, and early testing of its impact on the patient-doctor relationship. JMIR Mhealth Uhealth 2016 Mar 31;4(1):e24 [FREE Full text] [doi: 10.2196/mhealth.4546] [Medline: 27032541]
27. Spook JE, Paulussen T, Kok G, van Empelen P. Monitoring dietary intake and physical activity electronically: feasibility, usability, and ecological validity of a mobile-based Ecological Momentary Assessment tool. J Med Internet Res 2013 Sep 24;15(9):e214 [FREE Full text] [doi: 10.2196/jmir.2617] [Medline: 24067298]
28. Hong Y, Goldberg D, Dahlke DV, Ory MG, Cargill JS, Coughlin R, et al. Testing usability and acceptability of a web application to promote physical activity (iCanFit) among older adults. JMIR Hum Factors 2014 Oct 13;1(1):e2 [FREE Full text] [doi: 10.2196/humanfactors.3787] [Medline: 27025254]
29. Heinonen R, Luoto R, Lindfors P, Nygård CH. Usability and feasibility of mobile phone diaries in an experimental physical exercise study. Telemed J E Health 2012 Mar;18(2):115-119. [doi: 10.1089/tmj.2011.0087] [Medline: 22283356]
30. Kernot J, Olds T, Lewis LK, Maher C. Usability testing and piloting of the Mums Step It Up program--a team-based social networking physical activity intervention for women with young children. PLoS One 2014;9(10):e108842 [FREE Full text] [doi: 10.1371/journal.pone.0108842] [Medline: 25272039]
31. Kirwan M, Duncan MJ, Vandelanotte C, Mummery WK. Using smartphone technology to monitor physical activity in the 10,000 steps program: a matched case-control trial. J Med Internet Res 2012 Apr 20;14(2):e55 [FREE Full text] [doi: 10.2196/jmir.1950] [Medline: 22522112]
32. Naslund JA, Aschbrenner KA, Bartels SJ. Wearable devices and smartphones for activity tracking among people with serious mental illness. Ment Health Phys Act 2016 Mar;10:10-17 [FREE Full text] [doi: 10.1016/j.mhpa.2016.02.001] [Medline: 27134654]
33. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. J Strength Cond Res 2005 Feb;19(1):231-240. [doi: 10.1519/15184.1] [Medline: 15705040]
34. Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? Control Clin Trials 1996 Feb;17(1):1-12. [doi: 10.1016/0197-2456(95)00134-4] [Medline: 8721797]
35. Yates SL, Morley S, Eccleston C, de Williams AC. A scale for rating the quality of psychological trials for pain. Pain 2005 Oct;117(3):314-325. [doi: 10.1016/j.pain.2005.06.018] [Medline: 16154704]
36. Zapata BC, Fernández-Alemán JL, Idri A, Toval A. Empirical studies on usability of mHealth apps: a systematic literature review. J Med Syst 2015 Feb;39(2):1. [doi: 10.1007/s10916-014-0182-2] [Medline: 25600193]

37. Kelley K, Clark B, Brown V, Sitzia J. Good practice in the conduct and reporting of survey research. Int J Qual Health Care 2003 Jun;15(3):261-266. [doi: 10.1093/intqhc/mzg031] [Medline: 12803354]
38. Sousa VE, Dunn Lopez K. Towards usable e-Health. A systematic review of usability questionnaires. Appl Clin Inform 2017 May 10;8(2):470-490 [FREE Full text] [doi: 10.4338/ACI-2016-10-R-0170] [Medline: 28487932]
39. Lewis S, Warlow C. How to spot bias and other potential problems in randomised controlled trials. J Neurol Neurosurg Psychiatry 2004 Feb;75(2):181-187 [FREE Full text] [doi: 10.1136/jnnp.2003.025833] [Medline: 14742583]
40. Ellsworth MA, Dziadzko M, O'Horo JC, Farrell AM, Zhang J, Herasevich V. An appraisal of published usability evaluations of electronic health records via systematic review. J Am Med Inform Assoc 2017 Jan;24(1):218-226. [doi: 10.1093/jamia/ocw046] [Medline: 27107451]
41. Streiner DL, Norman GR, Cairney J. Health Measurement Scales: A Practical Guide to Their Development and Use. Fifth Edition. Oxford, UK: Oxford University Press; 2015.

Abbreviations

CAUSS: Critical Assessment of Usability Studies Scale
eHealth: electronic health
ICC: intraclass correlation coefficient
mHealth: mobile health

Edited by G Eysenbach; submitted 27.05.19; peer-reviewed by GE Iyawa, R Kok; comments to author 03.07.19; revised version received 26.07.19; accepted 13.08.19; published 15.11.19.

Please cite as:
Silva AG, Simões P, Santos R, Queirós A, Rocha NP, Rodrigues M
A Scale to Assess the Methodological Quality of Studies Assessing Usability of Electronic Health Products and Services: Delphi Study Followed by Validity and Reliability Testing
J Med Internet Res 2019;21(11):e14829
URL: http://www.jmir.org/2019/11/e14829/
doi:10.2196/14829
PMID:31730036

Original Paper

Quantifying Use of a Health Virtual Community of Practice for General Practitioners' Continuing Professional Development: A Novel Methodology and Pilot Evaluation

Abdulaziz Murad[1], MSc; Natalie Hyde[2], PhD; Shanton Chang[1], PhD; Reeva Lederman[1], PhD; Rachelle Bosua[1,3], PhD; Marie Pirotta[4], MBBS, PhD; Ralph Audehm[4], MBBS, DipRACOG; Christopher J Yates[5,6,7], MBBS, PhD; Andrew M Briggs[8], BSc Hons, PhD; Alexandra Gorelik[5,9], MSc; Cherie Chiang[5,6,7,10], MBBS, MD; John D Wark[5,6,7], MBBS, PhD

[1]School of Computing and Information Systems, University of Melbourne, Melbourne, Australia

[2]Deakin University, Geelong, Australia

[3]Open University of the Netherlands, Heerlen, Netherlands

[4]Department of General Practice, University of Melbourne, Melbourne, Australia

[5]Department of Medicine, Royal Melbourne Hospital, University of Melbourne, Melbourne, Australia

[6]Department of Diabetes and Endocrinology, Royal Melbourne Hospital, Melbourne, Australia

[7]Bone and Mineral Medicine, Royal Melbourne Hospital, Melbourne, Australia

[8]School of Physiotherapy and Exercise Science, Faculty of Health Sciences, Curtin University, Perth, Australia

[9]School of Behavioral and Health Science, Australian Catholic University, Melbourne, Australia

[10]Department of Pathology, Royal Melbourne Hospital, Melbourne, Australia

Corresponding Author:
John D Wark, MBBS, PhD
Department of Medicine
Royal Melbourne Hospital
University of Melbourne
The Royal Melbourne Hospital
Melbourne, 3050
Australia
Phone: 61 8344 3258
Email: jdwark@unimelb.edu.au

Abstract

Background: Health care practitioners (HPs), in particular general practitioners (GPs), are increasingly adopting Web-based social media platforms for continuing professional development (CPD). As GPs are restricted by time, distance, and demanding workloads, a health virtual community of practice (HVCoP) is an ideal solution to replace face-to-face CPD with Web-based CPD. However, barriers such as time and work schedules may limit participation in an HVCoP. Furthermore, it is difficult to gauge whether GPs engage actively or passively in HVCoP knowledge-acquisition for Web-based CPD, as GPs' competencies are usually measured with pre- and posttests.

Objective: This study investigated a method for measuring the engagement features needed for an HVCoP (the Community Fracture Capture [CFC] Learning Hub) for learning and knowledge sharing among GPs for their CPD activity.

Methods: A prototype CFC Learning Hub was developed using an Igloo Web-based social media software platform and involved a convenience sample of GPs interested in bone health topics. This Hub, a secure Web-based community site, included 2 key components: an online discussion forum and a knowledge repository (the Knowledge Hub). The discussion forum contained anonymized case studies (contributed by GP participants) and topical discussions (topics that were not case studies). Using 2 complementary tools (Google Analytics and Igloo Statistical Tool), we characterized individual participating GPs' engagement with the Hub. We measured the GP participants' behavior by quantifying the number of online sessions of the participants, activities undertaken within these online sessions, written posts made per learning topic, and their time spent per topic. We calculated time spent in both active and passive engagement for each topic.

Results: Seven GPs participated in the CFC Learning Hub HVCoP from September to November 2017. The complementary tools successfully captured the GP participants' engagement in the Hub. GPs were more active in topics in the discussion forum that had direct clinical application as opposed to didactic, evidence-based discussion topics (ie, topical discussions). From our knowledge hub, About Osteoporosis and Prevention were the most engaging topics, whereas shared decision making was the least active topic.

Conclusions: We showcased a novel complementary analysis method that allowed us to quantify the CFC Learning Hub's usage data into (1) sessions, (2) activities, (3) active or passive time spent, and (4) posts made to evaluate the potential engagement features needed for an HVCoP focused on GP participants' CPD process. Our design and evaluation methods for ongoing use and engagement in this Hub may be useful to evaluate future learning and knowledge-sharing projects for GPs and may allow for extension to other HPs' environments. However, owing to the limited number of GP participants in this study, we suggest that further research with a larger cohort should be performed to validate and extend these findings.

(J Med Internet Res 2019;21(11):e14545) doi:10.2196/14545

KEYWORDS

online systems; online social networking; general practitioners; online learning; continuing education; professional education; evaluation methodology; use-effectiveness; quantitative evaluation; knowledgebases; information sharing

Introduction

Background

Knowledgeable and skillful general practitioners (GPs) are fundamental for efficient and effective health care systems. In most communities, they are the primary source of health care to individuals and families [1,2], lending them the name *family practitioner/physician* [2,3]. Owing to the important role of GPs in the community, high-quality training and continuing professional development (CPD) are of utmost importance [2], creating an imperative to provide the necessary learning that GPs require to remain competent in their field [4,5]. GPs are required to continuously expand their knowledge and skills to ensure that evidence-based research is applied to provide the best care possible to patients [4,6]. Health care practitioners (HPs), GPs in particular, can find it challenging to maintain their CPD activities considering their heavy and often demanding workloads. Moreover, GPs and other HPs who practice in relatively isolated regions, which are common in a country such as Australia, can find it particularly difficult to participate in traditional educational activities. Therefore, new ways are required to enable and support efforts to maintain CPD activities.

GPs have become accustomed to using networking events (eg, conferences and out-of-hour lectures) and extensive reading to acquire new knowledge [7]. In addition, GPs seem to appreciate Web-based information to further develop practices and are increasingly seeking information online for their CPD [5,8-10]. GPs understand that Web-based CPD is a new, viable alternative to face-to-face learning (ie, conferences) that can be managed in their own time to further develop their learning and knowledge-sharing competencies for practice [11]. Hence, incorporating social media technologies for CPD has become a commonplace mechanism encouraging GPs' learning in online group settings (eg, Facebook and Twitter) [2,5,8].

A practitioner group working together on shared practices is defined as a community of practice. Health virtual communities of practice (HVCoPs) refer to a class of internet technologies used to share the best practices among HPs [8,12]. As a result of HVCoPs, Web-based community-based learning, sharing, and adopting of explicit evidence-based medical knowledge in work practices [8] by GPs [13] have arisen in the past decade.

Although HPs/GPs recognize the potential of using social media technologies for learning and knowledge sharing [8], they question whether using Web-based communities (eg, Facebook) to gain knowledge and share experiences for CPD is acceptable owing to privacy and trust issues [14,15]. Patient information being identifiable (eg, a rare disease that only a handful of patients have) to other online participants was considered a privacy concern. Another known privacy issue was GPs' personal information being identifiable (eg, identifying GPs by their location of practice) with constant online activity of sharing information. Not knowing if participants are real, practicing GPs or a random person online impersonating a GP was considered a trust issue. These concerns have lowered engagement over time [16]. GPs have also found it difficult to use other tools (eg, Web-based databases) to search for evidence-based research results, resulting in users and facilitators abandoning HVCoP systems [17]. Some GPs have also experienced difficulties in using mobile apps specifically developed to support literature searches [18]. GPs also differ in the type of CPD they require depending on the terms of appraisal needed or whether they are seeking to review existing knowledge or gain new knowledge [19]. Furthermore, GPs tend to lose interest over time even when fully engaged in an HVCoP (ie, user participation decreases after 2-3 months) as engagement for learning and knowledge-sharing activities in online forums is demanding on the participants' involvement (ie, users and facilitators) [13,20]. GPs often face other demands on their time that may limit participation [8,13]. As such, Web-based learning has not yet proven its benefits for learning, knowledge acquisition, and cost-effectiveness for GPs using Web-based systems [21]. Previous work (Multimedia Appendix 1) supported a set of required design principles to evaluate GPs' engagement through participation in our Community Fracture Capture (CFC) Learning Hub (an interactive, case-based, Web-based learning tool designed to help GPs improve the care of patients in relation to osteoporosis). It was hoped that such an HVCoP would help to mitigate barriers to participation in Web-based learning

[22-24]. Furthermore, active and passive GP engagement has not been formally studied in HVCoPs, as there has been no real way to track and quantify the participants' usage behavior. *Active users* post constantly in online discussion forums, whereas passive users merely engage in viewing content with no posting activity [13]. Previous studies typically [8,13,17,25] employed pre- and posttests to gauge whether GP participants, active or passive, acquired knowledge from the HVCoP to measure user engagement. Pre- and posttest evaluations for Web-based CPD are the most common forms of assessment, and there is a need for additional evaluation methods for GPs' learning outcomes [21]. Moreover, we are not aware of any previously described methods to measure GPs' engagement behaviors (ie, activity/usage) with Web-based learning CPD platforms at the individual level, let alone in the HVCoPs. Hence, to our knowledge, there are no current HVCoP platforms or studies that measure active and passive behavior to gauge GP participants' engagement.

Objectives

There is a need to better quantify GPs' participation in CPD through HVCoP usage, to establish whether active or passively engaged GPs' truly learn and acquire knowledge. In this paper, we report preliminary findings of GPs' learning and knowledge sharing in an HVCoP for CPD. In this ongoing research program to design and evaluate an HVCoP for GPs' CPD endeavors [22-24], we pose the research question: *What engagement features promote the use of an HVCoP for GPs' CPD?* This paper aimed to present the findings of a project undertaken to describe the performance features of a customized, interactive, case-based learning hub designed to help improve GPs' understanding and management of osteoporosis, a common condition that remains underdiagnosed and undertreated in many countries [26]. Hence, the main purpose of this study was the development and demonstration of a novel design and methodology for CPD and its evaluation.

Methods

The Prototype Platform

The CFC Learning Hub is a secure, Web-based prototype HVCoP website, created using the *Igloo* Web-based social media software platform. The CFC Learning Hub was developed for enhancing GPs' awareness and competence in caring for patients with osteoporosis. The project team comprised a mix of specialists in the field: experienced GPs; information systems researchers; technology experts; a project coordinator; and specialist physicians with expertise in bone health and osteoporosis, the CFC Learning Hub's theme. The project team, situated in Melbourne, Australia, drew on its learning, teaching, and clinical experience to define the following important design criteria for GP participants and elements in the project:

Practicing GPs was the target group, with potential inclusion of practice nurses (though there was some uncertainty whether this might inhibit contribution by some GPs).

Case-based learning preferably involving both experienced and trainee GPs' own case study contributions with anonymized patients (to encourage engagement).

Interactive engagement among all parties involved, that is, no didactic teaching component.

GP peers led the group as facilitators, with guidance where needed by specialist *advisers*, promoting case discussions and guiding discussion of content according to relevance and importance.

The total time commitment required for the CPD trial to make effective use of GPs' limited time, whereas the CFC Learning Hub format was to be flexible in terms of the timing of GP participants' input in the form of contributions (to more easily fit in with their workload commitments).

The CFC Learning Hub platform was developed with the assistance of an external developing entity (Involved—Design and Development Agency, Melbourne, Australia), which agreed that Igloo Web-based social media software was the best platform, kept development costs moderate, and accommodated the investigation team's design criteria. Our project team worked closely with Involved to develop and implement the CFC Learning Hub prototype with the support of 3 external GPs in a collaborative approach to the design of the HVCoP.

A facilitation team, formed to incentivize the engagement of the participants and facilitate discussions in the online discussion forum, included 4 specialist physicians, 2 senior GPs as facilitators, 1 dedicated content facilitator, and 1 information technology administrator. This team had moderation and administration rights throughout the CFC Learning Hub's life, whereas GP participants who joined and contributed case studies remained as participants throughout this period. GP participants were deidentified with an anonymous key for data management and analysis.

The CFC Learning Hub platform had 2 resources for GP interactions:

1. A Web-based knowledge repository (the Knowledge Hub) containing curated and prepopulated evidence-based research articles and other resources.
2. An online social network forum (the *Discussion Forum*) where GPs could freely post online comments, including:
   - Questions for discussion posted by facilitators;
   - Case studies to encourage GPs to learn and share their knowledge based on shared experiences and relevance to their immediate clinical practice;
   - Topical discussions as either (a) hot topics (HT) deemed relevant for GPs, posted by our osteoporosis specialists or (b) other topics (OT) that were open for wider discussion (ie, introductions, where facilitators and GPs introduce themselves, and burning questions, where GPs and facilitators post inquiries on osteoporosis).

The GP participants provided case studies as a requirement for joining the CFC Learning Hub. Facilitators and specialist advisers filtered and chose case studies that enabled the coverage of a syllabus of topics predetermined by members of the project team for this CPD course. Facilitators ensured that the posted case studies were anonymized and contained suitable, high-quality content for discussion. A total of 6 chosen case

studies were posted for discussion in the CFC Learning Hub at approximately weekly intervals.

The platform's topical discussions section included all OT posted in the discussion forum of the CFC Learning Hub. These included (1) an HT section chosen by the facilitators and specialist advisers to enrich GP participants' learning and knowledge sharing on issues identified as being of particular importance: diabetes and bone health, atypical femoral fracture, when to consider changing an osteoporosis therapy, and how to get the most from your patients' bone density testing; (2) an *Introduction* topic for all users to introduce themselves; and (3) a facility for GP participants to post inquiries based on seeking specific information about osteoporosis. Facilitators also could raise questions to promote discussion. These inquiries were organized under the term, *Burning Questions*. In terms of privacy, the CFC Learning Hub itself was a private network with a password log-in functionality that excluded online public entities (ie, people or organizations) outside of the HVCoP. In terms of trust, all GP participants and facilitators were known to each other within the CFC Learning Hub as all had customizable profiles and were not anonymized. We adopted this approach to instill trust among the participants as being genuine participants, consistent with standard Web-based private learning environments.

The knowledge hub acted as an accessible knowledge repository for all users at all times. This hub had 7 topics that the project team chose to include, each having detailed evidence-based research articles online. Furthermore, any new and interesting topics that the GP participants were discussing in the online discussion forum could be added at a later time by the facilitator team.

The lead time to the commencement of the CFC Learning Hub was 2.2 years (from December 2014 to February 2017). On average, the project lead (JDW) spent approximately 3 hours per week during the lead time to the project launch and 1.5 hours per week during the active phase of the project. The time spent by the investigation team was an average of 1 hour per project team member per fortnight leading to the launch of the CFC Learning Hub, and an average of 0.5 to 1 hour per week during the active phase of the project. The information systems doctoral student (AM) spent approximately 1 hour per week during the lead time to the project launch, and 3.5 hours per week during the active phase of the project (ie, tracking live interactions to update facilitators on engagement). The project coordinator (NH) was appointed at 0.4 full time equivalent during the lead time of the project. The developers (Involved—Design and Development Agency) had 5 meetings with the project team and proposed a work timeline to develop the CFC Learning Hub in 2 days, with time to be split between a producer, developer, and designer.

The direct project costs were Aus $170,000 (approximately US $114,000) and covered the following: Web developer fees, database management, Web hosting, content development, Igloo platform fees, ethics and governance submission fees, part-time study coordinator salary, GP facilitator consultancies, statistics support, and anticipated publication costs. In addition, AM received full-time support by a doctoral scholarship.

Upon the launch of the CFC Learning Hub, the time spent by facilitators depended on each facilitator, but there was an agreed expectation to be available once per day. Specialists involved from the facilitation team scheduled themselves with the assistance of the project coordinator, for each to be dedicated for a specific week of the CFC Learning Hub's full life cycle.

Figure 1 is a screenshot of the home screen and Figure 2 is an example of its discussion forum.

Figure 1. The Community Fracture Capture Learning Hub home screen.

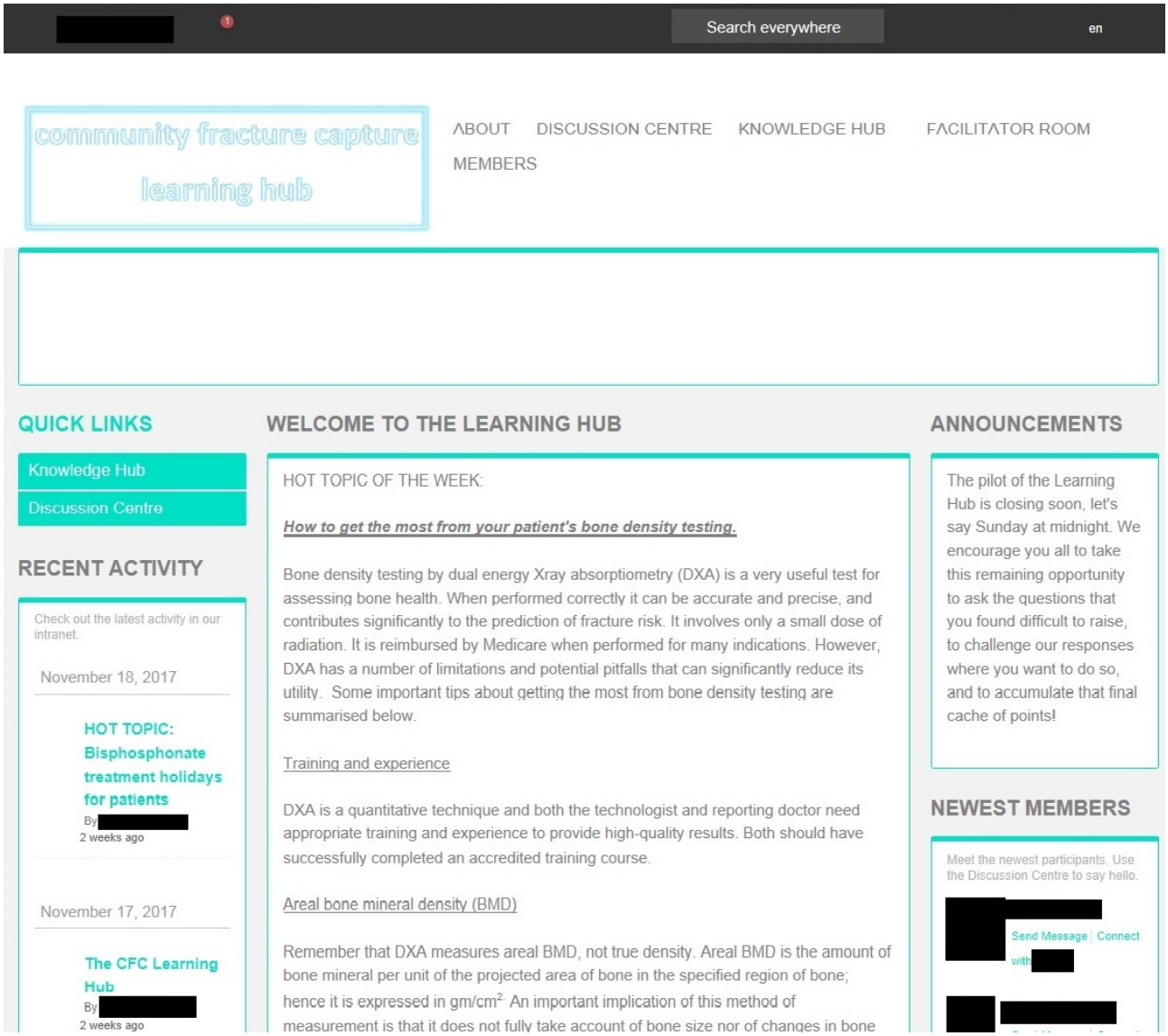
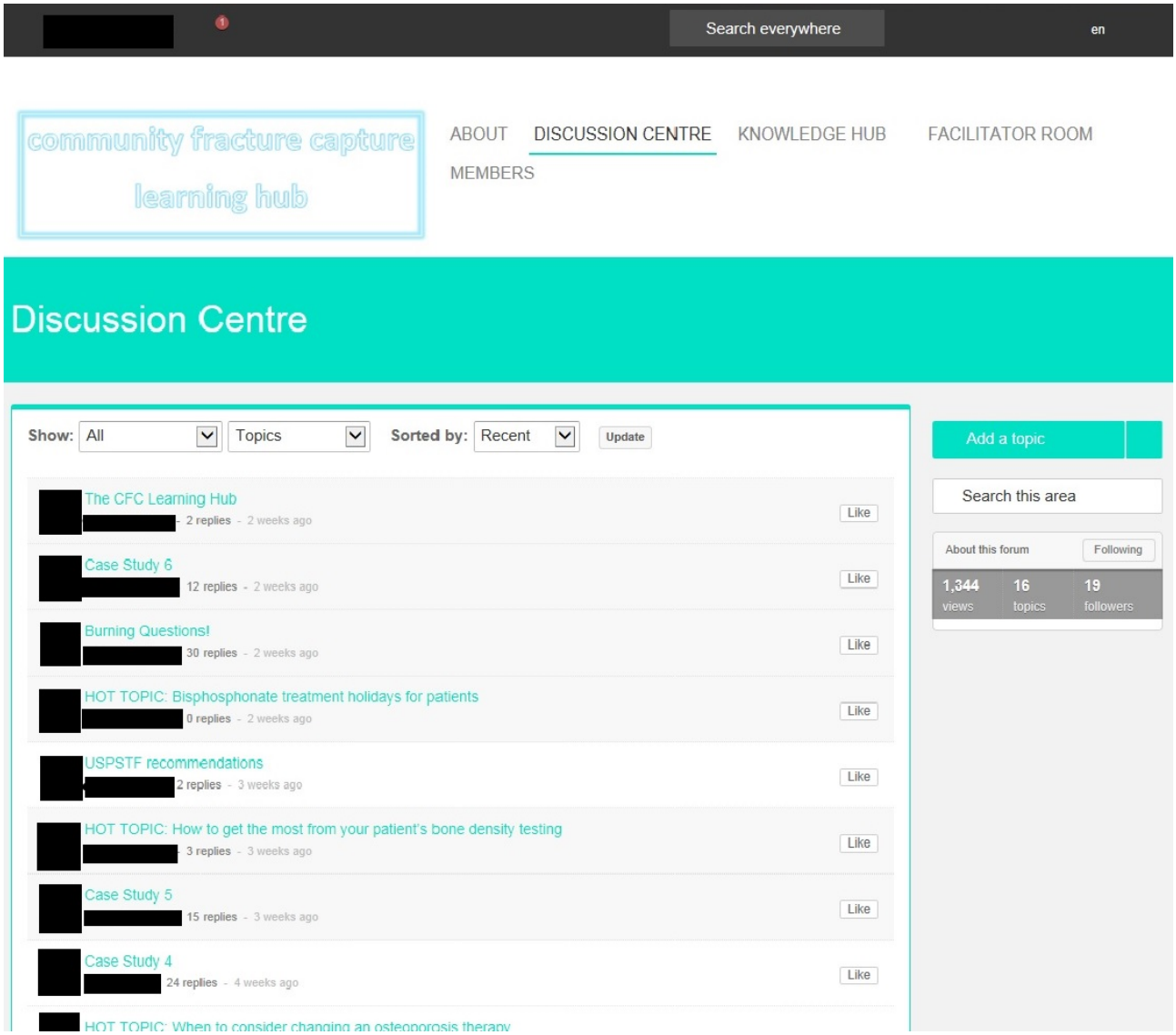


Figure 2. A brief graphical representation of using 2 analytical tools to process data.



Study Site

The study site was managed at the University of Melbourne and the Royal Melbourne Hospital. All users could sign up for password access to the CFC Learning Hub from one or more sites, for example, home or practice.

Data Storage and Security

All knowledge repository content was stored electronically on the Igloo data center platform. Each user had his/her own unique username and password to access the CFC Learning Hub. User data were also held in the Igloo data center in Toronto and Vancouver, Canada. GP participants were assigned a unique study number, and all collected data were deidentified (anonymized) and made reidentifiable only by linking separately stored password-protected GP participant information. All study data were password protected and accessible only with the approval of the study's principal investigator.

Study Timeline

Recruitment

The recruitment period was from February until August 2017.

Participation in Online Community Fracture Capture Learning Hub

The Web-based CPD course proceeded over 8 weeks (September to November 2017). Users received a 2-week introductory/familiarization period at the beginning of the course. Following this, the 6-week active period of the CFC Learning Hub commenced.

Study Population

Recruitment Procedure

The recruitment target was 15 GPs, identified from the GP participants of an earlier industry-sponsored osteoporosis education program and the Victorian Primary Care Practice-Based Research Network (VicReN) [27] mailing list. These GPs were sent a flyer inviting their participation in the project. Interested GPs contacted the project coordinator, after which they were screened for eligibility to participate in the program and subsequently enrolled if they satisfied the selection criteria and consented to participate.

Inclusion Criteria

GP participants were required to be a medical practitioner currently active in general practice in Australia, to submit preferably 2 or at least one suitable case study with discussion or learning points from their own clinical experience that included patients at risk or those with prevalent osteoporosis, and to have internet access.

Exclusion Criteria

GP participants excluded were those who were unable to provide patient case studies for discussion or were unable to commit to the anticipated time required to meet appropriate CPD guidelines (at least six hours).

Measures

Data related to sessions and activities of GP participants in the CFC Learning Hub were collected for all 8 weeks. Our focus was on the usage data to identify the GPs' learning and knowledge-sharing behaviors related to participation and engagement in the CFC Learning Hub. We defined the following terms from Google Analytics, henceforth termed GA:

- A session included all user-related activities from initial logging in to logging out of the CFC Learning Hub. All activities made between the logging in and logging out activities were grouped within a session.
- An activity included downloading of content, viewing posts and content, and posting comments.

The measurement of sessions and activities of a GP participant covered his or her entire active and passive engagement with the Web-based system, which could be used to measure the success of the prototype platform system (ie, high session count meant GPs were logging in, undertaking one or more activities, and then logging out). Actively engaged GPs were GP participants who posted a case study/topical discussion, whereas passively engaged GPs were GP participants who did not post but spent time only browsing a case study/topical discussion. In the knowledge hub, all GPs' participation was inherently passive as GPs were all browsing the knowledge hub database.

For measuring the use of the discussion forum, the usage behavior for each case study and topical discussion was assessed. For each case study and topical discussion, usage was measured from the creation date until the end of a 7-day period commencing from the creation date.

The use of the discussion forum was measured for the full CPD trial period (2 months) looking at the following:

- Each case study/topical discussion and its related discussion sessions for GP participants during the week after the case study/topical discussion was first published online. This showed the distribution of sessions among case studies/topical discussions and whether the GPs returned seeking more information, learning, and discussion. In addition, we assessed case studies/topical discussions that attracted the most activity by GPs.
- How long GP participants spent on each case study/topical discussion each week.
- The number of posts each GP made per case study/topical discussion per week. This measure also showed GPs who were actively and passively engaged per case study/topical discussion.

For the knowledge hub, we followed 2 key points in identifying engaging topics for GPs' CPD from a learning and knowledge-sharing perspective: (1) examining each knowledge hub's unique GP participant sessions and activities and (2) calculating how long GP participants spent on each knowledge hub topic.

Data Collection

Data were collected from 2 main sources: (1) the Igloo statistical tool, henceforth termed IGT, which can track the user's identification (eg, user 1 is John Doe) and (2) GA, which captures cumulative real-time use by users. However, in GA, users are not identifiable.

Both IGT and GA were used to collect and analyze the time that GP participants spent on each topic in the discussion forum and knowledge hub as follows:

- GA captured all users' (including facilitators and GP participants) sessions and activities in the CFC Learning Hub. However, all users were anonymous by default owing to GA's privacy terms and conditions. Hence, IGT was used to verify each activity, as IGT correctly identified each user ID in the CFC Learning Hub. This process was conducted manually to collate sessions and activities and avoid errors in identifying GP participants.
- IGT does not track how long each user spent on a given activity. However, GA tracked all activities and their duration for each user ID with time stamps. Hence, the 2 sets of results were cross-referenced with the time spent being calculated manually.

Although GA captures all facilitators' and GP participants' activities, GA does not differentiate between a view, download, or comment for each user, as it merges these together into 1 *Viewing* heading. IGT solves this issue by indicating each user's activity in detail but grouped together on a given period (eg, GP participant commented thrice on a given day). Hence, GA was relied upon to capture the activity of each GP participant, complemented by IGT to determine the type of activity for each individual, followed by cross-referencing with GA for the duration of activities. GA captured the cumulative real-time use of the participants. If a participant stopped using the platform for more than 30 min (the user opened the website page and became idle, ie, was not moving the mouse or was looking at another Web page), GA would automatically not count a session.

This method also measured whether GP participants were active (ie, posting comments to discuss) or passive (ie, viewing only). All data were collated in Microsoft Excel. See Multimedia Appendix 2.

Ethical Approval

The Melbourne Health Human Research Ethics Committee approved this project (site reference number: 2016.24). Electronic written consent from GP participants who joined the CFC Learning Hub was obtained to use their data. Electronic

written consent data were collected and managed by using Research Electronic Data Capture (REDCap) tools hosted at the University of Melbourne [28]. REDCap is a secure, Web-based application designed to support data capture for research studies, providing (1) an intuitive interface for validated data entry, (2) audit trails for tracking data manipulation and export procedures, (3) automated export procedures for seamless data downloads to common statistical packages, and (4) procedures for importing data from external sources [28]. The electronic consent also included GP waiver of consent of case study patients whose anonymized information was used in the project.

Results

Recruitment

A total of 19 GPs showed initial interest in joining the CFC Learning Hub. Sources of recruitment included VicReN, personal contacts from our GP facilitators, and from the Australian College of Rural and Remote Medicine website's CPD offerings. Of these 19 GPs, 8 committed to join. Later, 1 GP participant dropped out owing to family reasons, and 7 GPs continued for the full 8-week CPD course duration. Overall, 3 GP participants were above 50 years of age and 4 GP participants were aged under 50 years. Furthermore, 6 GP participants were females, and 1 was male. A total of 5 GP participants had more than 5 years of practice experience and 2 GP participants had less than 5 years of practice experience.

Case Studies

Using IGT and GA's complementary tools to collect and process data, we were able to identify, from over 90 unique IDs, 7 GP participants' use of various technologies (mobile, desktop, or tablet, given by GA specifically).

The case studies are described briefly in Table 1. In addition, this section highlights GP users' engagement in the case studies discussed in the discussion forum.

The total number of case study session engagement for all GP participants during their first week of creation:

- Case study 1 had 16.
- Case study 2 had 15.
- Case study 3 had 26.
- Case study 4 had 25.
- Case study 5 had 26.
- Case study 6 had 19.

The total number of sessions by all GP participants per case study topic ranged from 15 to 26, and the median number of GP participant sessions per topic was 22. We recorded 127 GP participant sessions in total.

The total number of topical discussion activities engagement for all GP participants during their first week of creation:

- Case study 1 had 17.
- Case study 2 had 20.
- Case study 3 had 31.
- Case study 4 had 31.
- Case study 5 had 32.
- Case study 6 had 26.

The total number of activities by all GP participants ranged from 17 to 32, and the median number of activities per topic was 28.5. In total there were 148 activities.

Table 2 presents the time spent by GP participants for each case study and the number of posts made for each given week, actively and passively. Table 2 presents a description of the engagement in each case study.

From Table 2, the total time spent per case study by all GP participants ranged from 32 min to 114 min, with a median of 86.5 min per case study. All GP participants put together spent a total of 458 min for all case studies.

The number of posts per case study ranged from 3 to 8, and the median number of GP participant posts was 5.5 for all case studies. The median number of active GP participants was 4 and the median number of passive GP participants was 2, for all case studies.

Table 1. Description of each case study's content.

Case studies	Description of case study content
Case study 1	A 59-year-old woman having bone mineral density measured as a health check
Case study 2	A 56-year-old woman with osteoporosis managed with raloxifene and physical activity
Case study 3	A 70-year-old woman with several osteoporosis risk factors
Case study 4	Osteoporosis in a 70-year-old man on androgen deprivation therapy postprostatectomy for prostate cancer
Case study 5	An 89-year-old woman with a history of vertebral fracture and previous osteoporosis therapy
Case study 6	A 67-year-old woman who has lost height and has dental problems

Table 2. Time spent (minutes) and posts made by all general practitioner participants on each case study during the first week of its creation. Active, passive, and nonengagements are included.

Case studies	Time spent in sessions by all GP[a] participants (min)	Posts made by all GP participants, n	GP participants actively engaged[b], n	GP participants passively engaged[c], n	GP participants not engaged, n
Case study 1	32	8	2	2	3
Case study 2	81	4	3	2	2
Case study 3	93	3	2	3	2
Case study 4	92	7	5	2	0
Case study 5	114	7	5	0	2
Case study 6	46	4	4	2	1
Total	458	33	N/A[d]	N/A	N/A

[a]GP: general practitioner.

[b]Participants who posted, browsed, and downloaded content.

[c]Participants who did not post but spent time browsing and downloading content.

[d]N/A: not applicable.

Topical Discussions

This section highlights the GP participants' engagement in topical discussions on the discussion forum of the CFC Learning Hub.

The total number of topical discussion session engagement for all GP participants during their first week of creation:

- Other topics 1 (OT1) Introduction had 3.
- Hot topic 1 (HT1): Diabetes and bone health had 8.
- Hot topic 2 (HT2): Atypical femoral fracture had 9
- Hot topic 3 (HT3): When to consider changing an osteoporosis therapy had 11.
- Hot topic 4 (HT4): How to get the most from your patients bone density testing had 6.
- Other topics 2 (OT2): Burning questions had 18.

The total number of sessions per topical discussion ranged from 3 to 18, and the median number of GP participant sessions per topical discussion was 8.5. The total number of sessions made by GP participants for all topical discussion was 55.

The total number of topical discussion activities engagement for all GP participants during their first week of creation:

- Other topics 1 (OT1) Introduction had 3.
- Hot topic 1 (HT1): Diabetes and bone health had 9.
- Hot topic 2 (HT2): Atypical femoral fracture had 9.
- Hot topic 3 (HT3): When to consider changing an osteoporosis therapy had 12.
- Hot topic 4 (HT4): How to get the most from your patients bone density testing had 7.
- Other topics 2 (OT2): Burning questions had 22.

The total number of activities per topical discussion ranged from 3 to 22, and the median per topic was 9. The total number of GP participant activities for all topical discussions was 57.

Table 3 presents the time spent by the GP participants on each topical discussion and the number of posts made for each given week, actively and passively.

In total, there were 14 burning questions covering 9 topics. The questions came from participating GPs and from GP facilitators. The topics were the assessment of an older patient with a recent peripheral fracture, the management of bone health in a young woman with an eating disorder, the selection of osteoporosis therapy, when to refer an osteoporosis patient to a specialist, the risks with osteoporosis therapy, what could GPs do better for their osteoporosis patients, the management of bone health in wheelchair-bound patients, the use of denosumab in patients with renal impairment, and the choice of therapy following a course of teriparatide.

As shown in Table 3, the total time spent by GP participants per topical discussion session ranged from 3 min to 72 min. The median time spent by all GPs for topical discussions was 19 min. GP participants spent a cumulative time of 189 min on topical discussions.

The post count by GP participants per topical discussion ranged from 0 to 10, and the median post count for all GP participants was 2. The total posts by GP participants for all topical discussions were 16. However, it should be noted that there were 14 burning questions; therefore, the number of posts for burning questions cannot be directly compared with those for HT. The median number of actively engaged GP participants was 1 and the median number of passively engaged GPs was 3, for all topical discussions.

Table 3. Total time spent (minutes) and posts made by general practitioner participants on each topical discussion during the first week of its creation. Active, passive, and nonengagements are included.

Topical discussions	Time spent in sessions by all GP[a] participants (min)	Posts made by all GP participants, n	GP participants actively engaged, n	GP participants passively engaged, n	GP participants not engaged, n
Other topic 1: Introductions	3	0	0	3	4
Hot topic 1: Diabetes and bone health	10	1	1	3	3
Hot topic 2: Atypical femoral fracture	61	3	2	2	3
Hot topic 3: When to consider changing an osteoporosis therapy	24	1	1	4	2
Hot topic 4: How to get the most from your patient's bone density testing	19	1	1	4	2
Other topic 2: Burning questions[b]	72	10	2	2	3
Total	189	16	NA[c]	NA	NA

[a]GP: general practitioner.

[b]This category contained 14 questions and 9 topics.

[c]N/A: Not applicable.

Knowledge Hub

This section includes GP participants who were engaged for sessions and activities in the knowledge hub from the CFC Learning Hub.

The total number of Knowledge Hub sessions engagement by all GP participants throughout the trial:

- KH1: About osteoporosis had 10.
- KH2: Diagnosis had 5.
- KH3: Patient resources had 6.
- KH4: Prevention had 8.
- KH5: Risk assessment had 10.
- KH6: Shared decision making had 0.
- KH7: Treatment had 9.

The number of sessions per knowledge hub topic ranged from 0 to 10, and the median per knowledge hub topic by all GP participants was 8. In total, GP participants undertook 48 knowledge hub sessions.

The total number of Knowledge Hub activities engagement by all GP participants throughout the trial:

- KH1: About osteoporosis had 12.
- KH2: Diagnosis had 6.
- KH3: Patient resources had 9.
- KH4: Prevention had 9.
- KH5: Risk assessment had 14.
- KH6: Shared decision making had 0.
- KH7: Treatment had 14.

The number of GP participants' activities per knowledge hub topic ranged from 0 to 14, with a total of 64 activities and a median of 9 activities per knowledge hub topic.

Table 4 presents the total time spent on each knowledge hub topic in sessions by all GP participants and the number of GP participants who engaged in each knowledge hub topic over the duration of the trial.

As shown in Table 4, the time spent by GP participants in the knowledge hub ranged from 0 min to 226 min per knowledge hub topic, with the median time spent being 152 min. The total time spent by GP participants was 1057 min. The number of GP participants who engaged in each topic ranged from 0 to 6. The median number who engaged per topic was 4 and the median number of nonengaged participants per topic was 3.

Table 4. Time spent (minutes) by general practitioner participants in sessions and the number who engaged in each knowledge hub topic.

KH[a] topic	Total time spent by all GP[b] participants on KH topics (min)	GP participants engaged in each KH topic[c], n	GP participants not engaged in each KH topic, n
KH1: About osteoporosis	223	6	1
KH2: Diagnosis	143	4	3
KH3: Patient resources	140	3	4
KH4: Prevention	226	6	1
KH5: Risk assessment	152	4	3
KH6: Shared decision making	0	0	7
KH7: Treatment	173	4	3
Total	1057	N/A[d]	N/A

[a]KH: knowledge hub.

[b]GP: general practitioner.

[c]Engagement in knowledge hub topics was passive only.

[d]N/A: not applicable.

Discussion

Principal Findings

The objective of this study was to test the methods we developed to quantify GP participants' active and passive interactions within an HVCoP platform, which we designed as a learning tool for patient care in osteoporosis. Furthermore, to our knowledge no other HVCoP reported in the literature had the ability to capture individual GPs' engagement in detail, which our study managed to do, and to explain the intricacies of the method itself. The key features of the platform were that the material presented was centered around the clinical cases provided by the participating GPs themselves and that the learning activities were designed to be interactive. In this functional HVCoP platform, 2 tools (ie, GA and IGT) were combined to capture (1) the time spent by each GP participant, (2) posts made by each GP participant, and (3) the specific activities performed within these sessions to verify GP participants' active, passive, and nonuse exactly. The platform and the associated analytical capability appear to have several characteristics suitable for both formal and informal CPD activities and may provide an attractive, cost-effective approach to CPD for busy health care professionals, particularly those practicing in rural and regional locations, with the incentive of receiving CPD points.

The novelty of this study centered around the ability to capture the GP participants' engagement in discussing their own case studies, curated by facilitators, inside a secure and private HVCoP platform for a real-world medical problem (ie, osteoporosis).

The sample of 7 GP participants was small, but it might be speculated from their active and passive behavior that they preferred *practice-based* topics (ie, case studies) rather than didactic information on osteoporosis (ie, topical discussions). An example would be our most engaging case study, case study 5, where GP participants and facilitators discussed an 89-year-old woman who had a history of vertebral fracture noting previous osteoporosis therapy. Previous literature suggests that case studies are important for incentivizing HPs in using Web-based social networks [29,30], especially *practice-based* topics by GPs [13]. This study supports both these areas of previous work; however, given the limitation of our small sample size, this notion still needs to be tested and verified in a future larger study to truly understand and explain this behavior.

The knowledge hub topics that engaged the most GP participants (6 out of 7) were *About Osteoporosis* and *Prevention*. The *About Osteoporosis* topic presented general information about osteoporosis and linked resources for further knowledge acquisition. *Prevention* is knowledge about the prevention of osteoporosis. GPs tend to benchmark their knowledge on specific medical conditions [13,31]. Our study supported this notion as GPs engaged in benchmarking their own knowledge with the *About Osteoporosis* topic. Furthermore, previous research highlights that GPs are more focused on learning about preventive methods for osteoporosis and not necessarily treatment as research in treating osteoporosis continues to address important unanswered questions [32,33]; hence, this may explain the GP participants' engagement in our *Prevention* topic. Our least active knowledge hub topic concerned shared decision making. This observation was based on the finding that no participants engaged with this topic, which seems to contradict the literature about the notion of patient-physician shared decision making. However, many of those studies looked at the patients' perception on the matter [34,35], and HPs have expressed doubts about the very notion of shared decision making at its core [34,36]. With our limited sample size, we hypothesized that GP participants may have believed that they were familiar with this topic in their practice and did not need to see/learn more about it from the HVCoP. In addition, GPs might be accustomed to this concept across different medical conditions in their practice already and the intention of joining the CFC Learning Hub was for specific clinical skills and knowledge related to a specific medical condition. Hence, they may have been less willing to devote time to a theme based on overall practice in osteoporosis and communication behavior.

A larger study is required to properly explore the above speculations from our study results. Nevertheless, the results to date strongly suggest that the platform we have developed will be capable of collecting the necessary quantifiable information for analysis in a future larger study and that this platform will prove to be a useful vehicle for health care professionals' CPD activities.

Limitations and Future Research

This study has several limitations. As mentioned above, the sample was small. On the contrary, the investigators considered this sample sufficient for the study program to test our measurement and analysis method and GP participants were very active throughout. This study assessed engagement in relation to a single clinical condition (ie, osteoporosis). Different GPs may have an interest in different topics and therefore measuring engagement in relation to other medical topics also needs to be explored in future research to overcome any selection bias associated with sampling in this study. Studies of other HVCoPs indicate that many had an induction session before commencement. GP participants could attend an informal discussion at the study site, both pre- and postparticipation in the program (ie, in a study by Barnett et al [13]). For administrative and logistical reasons, this study did not have such a session which might have encouraged engagement. Instead, we used the first 2 weeks for GP participants to introduce themselves as well as case study 1 and HT 1 to become familiar with the CFC Learning Hub. The study was implemented only in Melbourne, Australia, and the platform might be limited in its application to other contexts (eg, geographical regions with a paucity of experienced specialist advisers or where GPs were uninterested in particular medical conditions). A total of 6 out of the 7 GP participants were females. Hence, we cannot rule out a gender difference in engagement with this CPD tool as designed and tested. A larger sample with a balanced gender ratio should be evaluated in future studies. Furthermore, the analysis was made by 1 researcher following the method described in the Data Collection section. However, data that were gathered in Microsoft Excel were checked with a specialist and the project coordinator for validity. In addition, the approach that we tested to characterize and quantify participation can be automated and this modification should be feasible for future work.

Therefore, although there are some strong insights from our results, further investigation of the perspectives of the GP participants who were involved will also enhance our understanding of their use of the HVCoP. An example would be to compare our results measured here regarding engagement with our pre- and postknowledge testing to triangulate whether our HVCoP for CPD had positive, neutral, or negative knowledge outcomes for GP participants. Another example is qualitatively assessing GP participants in a postuse interview session about their engagement experiences in the CFC Learning Hub. This will be examined in forthcoming publications.

Conclusions

This study presents a method to quantify GP participants' hub-related sessions, activities in each session, time spent, and posting behavior as evidence of what potential features can encourage GP participants to engage in learning and knowledge sharing in an HVCoP for CPD. Furthermore, our study suggests new avenues of CPD by providing evidence of learning and knowledge acquisition outside traditional authoritative sources. The study also suggests new ways of tracking the path from engagement to behavioral use in receiving accreditation (ie, CPD points).

Compared with other Web-based educational communities, perhaps our main insight is that GP participants are interested and engaged in practice-based Web-based learning where they can discuss cases with other professionals. This challenges the cost-effectiveness of building large websites with significant libraries of materials to support Web-based learning. Combining the 2 analytical tools made it possible to measure the duration of time spent and the specific activities performed within these sessions to measure the GP participants' use of the learning hub.

This study can inform a larger study by focusing on creating and facilitating more practice-based topics with GP participants. Furthermore, we propose posting a query topic such as our *Burning Questions* to give an opportunity for GPs to ask general osteoporosis questions not covered in curated case study posts. This query topic can be posted once every 3 to 4 case studies and might help in incentivizing GPs to be engaged with the HVCoP as a means of benchmarking their own current knowledge with the specialists involved.

Acknowledgments

The authors thank Dr Sarah Price for her valuable assistance as a specialist adviser in the CFC Learning Hub. The authors also thank Ms Ornella Clavisi for her consumer perspective on GPs learning online about osteoporosis. Author AMB was supported by a fellowship awarded by the Australian National Health and Medical Research Council, number 1132548. Author AM was funded and supported by a doctoral scholarship provided by the Ministry of Education, Saudi Arabia. Amgen and Actavis-Allergan provided funding for this study, none of whom had any role in study design, data collection nor interpretation, nor drafting of this manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

A brief explanation of Community Fracture Capture Hub design principles from our previous work.
[PDF File (Adobe PDF File), 74 KB - jmir_v21i11e14545_app1.pdf]

Multimedia Appendix 2
A brief graphical representation of using two analytical tools to process data.
[PDF File (Adobe PDF File), 147 KB - jmir_v21i11e14545_app2.pdf]

References

1. Starfield B, Shi L, Macinko J. Contribution of primary care to health systems and health. Milbank Q 2005;83(3):457-502 [FREE Full text] [doi: 10.1111/j.1468-0009.2005.00409.x] [Medline: 16202000]
2. Barnett S, Jones SC, Bennett S, Iverson D, Bonney A. General practice training and virtual communities of practice - a review of the literature. BMC Fam Pract 2012 Aug 21;13:87 [FREE Full text] [doi: 10.1186/1471-2296-13-87] [Medline: 22905827]
3. Olesen F, Dickinson J, Hjortdahl P. General practice--time for a new definition. Br Med J 2000 Feb 5;320(7231):354-357 [FREE Full text] [doi: 10.1136/bmj.320.7231.354] [Medline: 10657333]
4. Norman GR, Shannon SI, Marrin ML. The need for needs assessment in continuing medical education. Br Med J 2004 Apr 24;328(7446):999-1001 [FREE Full text] [doi: 10.1136/bmj.328.7446.999] [Medline: 15105326]
5. VanNieuwenborg L, Goossens M, De Lepeleire J, Schoenmakers B. Continuing medical education for general practitioners: a practice format. Postgrad Med J 2016 Apr;92(1086):217-222 [FREE Full text] [doi: 10.1136/postgradmedj-2015-133662] [Medline: 26850504]
6. Zarif Sanaiey N, Karamnejad S, Rezaee R. Educational needs of family physicians in the domains of health and conformity with continuing education in Fasa University of Medical Sciences. J Adv Med Educ Prof 2015 Apr;3(2):84-89 [FREE Full text] [Medline: 25927073]
7. Mitchell G, Cherry M, Kennedy R, Weeden K, Burridge L, Clavarino A, et al. General practitioner, specialist providers case conferences in palliative care--lessons learned from 56 case conferences. Aust Fam Physician 2005 May;34(5):389-392 [FREE Full text] [Medline: 15887946]
8. Rolls K, Hansen M, Jackson D, Elliott D. How health care professionals use social media to create virtual communities: an integrative review. J Med Internet Res 2016 Jun 16;18(6):e166 [FREE Full text] [doi: 10.2196/jmir.5312] [Medline: 27328967]
9. Yee M, Simpson-Young V, Paton R, Zuo Y. How do GPs want to learn in the digital era? Aust Fam Physician 2014 Jun;43(6):399-402 [FREE Full text] [Medline: 24897992]
10. Fan H, Smith S, Lederman R, Chang S. Why People Trust in Online Health Communities: An Integrated Approach. In: Proceedings of the 21st Australasian Conference on Information Systems. 2010 Presented at: ACIS'10; December 1-3, 2010; Brisbane URL: https://ssrn.com/abstract=1775303
11. Bjerre LM, Paterson NR, McGowan J, Hogg W, Campbell C, Viner G, et al. Do Continuing Medical Education (CME) events cover the content physicians want to know? A content analysis of CME offerings. J Contin Educ Health Prof 2015;35(1):27-37. [doi: 10.1002/chp.21268] [Medline: 25799970]
12. Wenger E, McDermott R, Snyder W. Cultivating Communities of Practice: A Guide to Managing Knowledge. Brighton, Massachusetts: Harvard Business Press; 2002.
13. Barnett S, Jones SC, Caton T, Iverson D, Bennett S, Robinson L. Implementing a virtual community of practice for family physician training: a mixed-methods case study. J Med Internet Res 2014 Mar 12;16(3):e83 [FREE Full text] [doi: 10.2196/jmir.3083] [Medline: 24622292]
14. Maisonneuve H, Chambe J, Lorenzo M, Pelaccia T. How do general practice residents use social networking sites in asynchronous distance learning? BMC Med Educ 2015 Sep 21;15:154 [FREE Full text] [doi: 10.1186/s12909-015-0435-x] [Medline: 26391989]
15. Lederman R, Shanks G, Gibbs M. Meeting Privacy Obligations: The Implications for Information Systems Development. In: Proceedings of the European Conference on Information Systems. 2003 Presented at: ECIS'03; June 16-21, 2003; Naples, Italy URL: https://aisel.aisnet.org/ecis2003/96
16. MacWalter G, McKay J, Bowie P. Utilisation of internet resources for continuing professional development: a cross-sectional survey of general practitioners in Scotland. BMC Med Educ 2016 Jan 21;16:24 [FREE Full text] [doi: 10.1186/s12909-016-0540-5] [Medline: 26791566]
17. Fan S, Cooling N, Radford J, Fabian D, Brown J. Building on the evaluation of STARS: using online repositories to support the general practice learning community. Eval J Aust 2014;14(2):25-41. [doi: 10.1177/1035719x1401400204]
18. Sindi S, Calov E, Fokkens J, Ngandu T, Soininen H, Tuomilehto J, et al. The CAIDE Dementia Risk Score App: The development of an evidence-based mobile application to predict the risk of dementia. Alzheimers Dement (Amst) 2015 Sep;1(3):328-333 [FREE Full text] [doi: 10.1016/j.dadm.2015.06.005] [Medline: 27239514]

19. Salinas GD. Trends in physician preferences for and use of sources of medical information in response to questions arising at the point of care: 2009-2013. J Contin Educ Health Prof 2014;34(Suppl 1):S11-S16. [doi: 10.1002/chp.21224] [Medline: 24935878]
20. Preece J. Sociability and usability in online communities: determining and measuring success. Behav Inform Tech 2010 Nov 08;20(5):347-356. [doi: 10.1080/01449290110084683]
21. George PP, Zhabenko O, Kyaw BM, Antoniou P, Posadzki P, Saxena N, et al. Online digital education for postregistration training of medical doctors: systematic review by the digital health education collaboration. J Med Internet Res 2019 Feb 25;21(2):e13269 [FREE Full text] [doi: 10.2196/13269] [Medline: 30801252]
22. Murad A, Bosua R, Chang S, Lederman R, Wark JD. Informing the Design of a Health Virtual Community of Practice: A Knowledge Translation Approach. In: Proceedings of the 25th European Conference on Information Systems. 2018 Presented at: ECIS'18; June 23-28, 2018; Portsmouth, United Kingdom URL: https://aisel.aisnet.org/ecis2018_rip/24
23. Murad A, Lederman R, Bosua R, Chang S, Pirotta M, Wark JD. Health Practitioner Perceptions of using a Health Virtual Community of Practice for their Continuing Professional Development. In: Proceedings of the 21st Pacific Asia Conference on Information Systems. 2017 Presented at: PACIS'17; July 16-20, 2017; Langkawi, Malaysia URL: https://aisel.aisnet.org/pacis2017/227
24. Murad A, Lederman R, Bosua R, Chang S, Wark JD. Enhancing general practitioners participation in a virtual community of practice for continuing medical education: an exploratory study. Stud Health Technol Inform 2017;239:97-103. [doi: 10.3233/978-1-61499-783-2-97] [Medline: 28756443]
25. Pratte G, Hurtubise K, Rivard L, Berbari J, Camden C. Developing a web platform to support a community of practice: a mixed methods study in pediatric physiotherapy. J Contin Educ Health Prof 2018;38(1):19-24. [doi: 10.1097/CEH.0000000000000185] [Medline: 29369269]
26. The Lancet Diabetes Endocrinology. Osteoporosis: a roadmap to close the treatment gap. Lancet Diabetes Endocrinol 2018 Nov;6(11):833. [doi: 10.1016/S2213-8587(18)30292-4] [Medline: 30322720]
27. Soós M, Temple-Smith M, Gunn J, Johnston-Ata'Ata K, Pirotta M. Establishing the victorian primary care practice based research network. Aust Fam Physician 2010 Nov;39(11):857-862 [FREE Full text] [Medline: 21301660]
28. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
29. Falkman G, Gustafsson M, Jontell M, Torgersson O. SOMWeb: a semantic web-based system for supporting collaboration of distributed medical communities of practice. J Med Internet Res 2008 Aug 26;10(3):e25 [FREE Full text] [doi: 10.2196/jmir.1059] [Medline: 18725355]
30. Sharma S, Smith AF, Rooksby J, Gerry B. Involving users in the design of a system for sharing lessons from adverse incidents in anaesthesia. Anaesthesia 2006 Apr;61(4):350-354 [FREE Full text] [doi: 10.1111/j.1365-2044.2006.04542.x] [Medline: 16548954]
31. Lara B, Cañas F, Vidal A, Nadal N, Rius F, Paredes E, et al. Knowledge management through two virtual communities of practice (Endobloc and Pneumobloc). Health Informatics J 2017 Sep;23(3):170-180. [doi: 10.1177/1460458216639739] [Medline: 27102887]
32. Ganda K, Puech M, Chen JS, Speerin R, Bleasel J, Center JR, et al. Models of care for the secondary prevention of osteoporotic fractures: a systematic review and meta-analysis. Osteoporos Int 2013 Feb;24(2):393-406. [doi: 10.1007/s00198-012-2090-y] [Medline: 22829395]
33. Gérvas J, Starfield B, Heath I. Is clinical prevention better than cure? Lancet 2008 Dec 6;372(9654):1997-1999. [doi: 10.1016/S0140-6736(08)61843-7] [Medline: 19059049]
34. Elwyn G, Frosch D, Thomson R, Joseph-Williams N, Lloyd A, Kinnersley P, et al. Shared decision making: a model for clinical practice. J Gen Intern Med 2012 Oct;27(10):1361-1367 [FREE Full text] [doi: 10.1007/s11606-012-2077-6] [Medline: 22618581]
35. Barry MJ, Edgman-Levitan S. Shared decision making--pinnacle of patient-centered care. N Engl J Med 2012 Mar 1;366(9):780-781. [doi: 10.1056/NEJMp1109283] [Medline: 22375967]
36. Coulter A. Do patients want a choice and does it work? Br Med J 2010 Oct 14;341:c4989. [doi: 10.1136/bmj.c4989] [Medline: 20947576]

Abbreviations

CFC: Community Fracture Capture
CPD: continuing professional development
GA: Google Analytics
GP: general practitioner
HP: health care practitioner
HT: hot topics
HVCoP: health virtual community of practice

IGT: igloo statistical tool
OT: other topics
REDCap: Research Electronic Data Capture
VicReN: Victorian Primary Care Practice-Based Research Network

Edited by G Eysenbach; submitted 02.05.19; peer-reviewed by S Pit, T Ndabu; comments to author 26.07.19; revised version received 18.09.19; accepted 24.09.19; published 27.11.19.

Please cite as:
Murad A, Hyde N, Chang S, Lederman R, Bosua R, Pirotta M, Audehm R, Yates CJ, Briggs AM, Gorelik A, Chiang C, Wark JD Quantifying Use of a Health Virtual Community of Practice for General Practitioners' Continuing Professional Development: A Novel Methodology and Pilot Evaluation
J Med Internet Res 2019;21(11):e14545
URL: http://www.jmir.org/2019/11/e14545/
doi:10.2196/14545
PMID:31774401

Original Paper

Medical Students' Experiences and Outcomes Using a Virtual Human Simulation to Improve Communication Skills: Mixed Methods Study

Timothy C Guetterman[1], PhD; Rae Sakakibara[2], BA; Srikar Baireddy[2]; Frederick W Kron[2], MD; Mark W Scerbo[3], PhD; James F Cleary[4], MD; Michael D Fetters[2], MD, MPH, MA

[1]Creighton University, Omaha, NE, United States
[2]University of Michigan, Ann Arbor, MI, United States
[3]Old Dominion University, Norfolk, VA, United States
[4]Indiana University, Indianapolis, IN, United States

Corresponding Author:
Timothy C Guetterman, PhD
Creighton University
2500 California Plaza
Omaha, NE, 68178
United States
Phone: 1 402 280 4778
Email: timguetterman@creighton.edu

Abstract

Background: Attending to the wide range of communication behaviors that convey empathy is an important but often underemphasized concept to reduce errors in care, improve patient satisfaction, and improve cancer patient outcomes. A virtual human (VH)–based simulation, MPathic-VR, was developed to train health care providers in empathic communication with patients and in interprofessional settings and evaluated through a randomized controlled trial.

Objective: This mixed methods study aimed to investigate the differential effects of a VH-based simulation developed to train health care providers in empathic patient-provider and interprofessional communication.

Methods: We employed a mixed methods intervention design, involving a comparison of 2 quantitative measures—MPathic-VR–calculated scores and the objective structured clinical exam (OSCE) scores—with qualitative reflections by medical students about their experiences. This paper is a secondary, focused analysis of intervention arm data from the larger trial. Students at 3 medical schools in the United States (n=206) received simulation to improve empathic communication skills. We conducted analysis of variance, thematic text analysis, and merging mixed methods analysis.

Results: OSCE scores were significantly improved for learners in the intervention group (mean 0.806, SD 0.201) compared with the control group (mean 0.752, SD 0.198; $F_{1,414}$=6.09; P=.01). Qualitative analysis revealed 3 major positive themes for the MPathic-VR group learners: gaining useful communication skills, learning awareness of nonverbal skills in addition to verbal skills, and feeling motivated to learn more about communication. Finally, the results of the mixed methods analysis indicated that most of the variation between high, middle, and lower performers was noted about nonverbal behaviors. Medium and high OSCE scorers most often commented on the importance of nonverbal communication. Themes of motivation to learn about communication were only present in middle and high scorers.

Conclusions: VHs are a promising strategy for improving empathic communication in health care. Higher performers seemed most engaged to learn, particularly nonverbal skills.

(J Med Internet Res 2019;21(11):e15459) doi:10.2196/15459

KEYWORDS

cancer; virtual reality; health communication; interprofessional relations; informatics; nonverbal communication; computer simulation; physician-nurse relations; empathy

Introduction

Background

Communication is critical to health encounters. Poor communication in health care has been linked to adverse consequences, medical errors, and decreased satisfaction [1-6]. Both patient-provider and provider-provider communication skills contribute to outcomes. Communication is foundational to many aspects of the health encounter from imparting information and diagnoses to offering explanations for medical decision making. The quality of those explanations (ie, *causability*) relies on high-quality communication [7]. In addition to imparting and gathering information, the manner in which providers communicate is equally important and often underemphasized. To address this critical need, MPathic-VR is an intervention that leverages virtual human (VH) informatics technology to target training empathic communication skills. Empathy is a set of constructs that relate to the "response of one individual to the experiences of another" [8]. In contrast, sympathy is typically a reference to feeling of compassion for others. Constructs of empathy include the following: (1) antecedents—individuals and situations, (2) processes—mechanisms to produce an empathic outcome, (3) intrapersonal empathic outcomes—cognitive and emotional responses within the observer, and (4) interpersonal empathic outcomes—behavioral responses of the observer directed toward the target [8]. These definitions assume 2 or more individuals, an observer (eg, a provider) who is in a situation and is responding empathically and a target individual (eg, a patient) with whom the observer interacts.

Teaching empathy includes both cognitive and emotional domains [9,10]. "Empathy is the foundation of patient care" [10] that is cultivated through actively listening to patients. Despite its foundational importance, training in empathy receives relatively little attention and even erodes during medical school [11].

Communicating with empathy requires attention to both verbal and nonverbal communication, such as microencouragers, proximity, eye contact, nodding, and appropriate use of smiles. Interventions with providers tend to give skills training in a variety of formats, such as videos, courses, or workshops [12,13]. Computer-based conversational agents have been developed to address cognitive tasks [14] such as verbal communication [15,16] and reasoning for diagnosis and therapy. A useful conversational agent is VHs, which have human appearance and the ability to interact, responding to humans and engaging in communication behaviors as in a typical conversation [17].

The body of VH literature related to health communication training is relatively small. Yet, research on VH to enhance health communication shows promise in training medical students and nurses [15,18-21]. VHs offer a unique advantage by providing an authentic yet low-risk simulated environment to learn with the appropriate level of challenge [22]. Users perceive VH interactions as real [23] social situations, and they perceive that authenticity can enhance learning and engagement [22,24,25]. The underlying informatics system can direct the learner through an adaptive path through a scenario based on responses. These systems collect data (eg, verbal responses) as the learner interacts with the VH, process the interaction, and provide real-time automated feedback to the learner. After receiving formative feedback, the learner can immediately reflect and repeat the simulation experience, which changes depending on the learner's actions. Reflection and deliberate practice are crucial to developing communication skills [26-29].

In contrast to human standardized patients, by leveraging informatics technology, VHs present information in a consistent manner, eliminating the variability of repeated human performances without becoming fatigued. As an informatics-based system, VHs could interpret nuances of nonverbal behavior and facial expression dynamically and then present its formative feedback to learners in the after-action review, eliminating the need for labor-intensive video analysis and conversation analysis.

Virtual Human Communication Simulation

In a parent study, a multisite single-blinded mixed methods randomized controlled trial [30], medical students were randomly assigned to 1 of the 2 conditions: an MPathic-VR intervention or a computer-based learning control module [30]. MPathic-VR is an informatics-based technology that engages learners in a conversation with VH characters to provide training in both verbal and nonverbal communication. Using active learning strategies, the simulation is designed for learners to have authentic, challenging conversations with VHs. The simulation uses intelligent VHs that detect body motion, facial expressions, and speech. The system consists of a computer with a widescreen monitor, a microphone to recognize speech, and a Microsoft Kinect sensor to detect predefined nonverbal communication behaviors (ie, smiles, nodding, body leaning, and eyebrow raises). MPathic-VR trains the learner in nonverbal behaviors, which the user must demonstrate to the system as picked up by a sensor before continuing.

The intervention involved interactive modules with 3 VHs: a young Latina woman who developed leukemia, her mother, and a nurse providing care for the woman. The learner participated in 2 scenarios—intercultural and interprofessional—that required expressing empathy and demonstrating nonverbal listening skills. At points needing communication from the learner, each scenario paused, and the system presented 3 possible responses. The system recognized the response the learner provided to the VH, which led to a commensurate path in the scenario (eg, a poor response would escalate the situation). After completing each full scenario, the system provided automated feedback, and the learner repeated the scenario. Medical students randomized to the intervention were assigned an MPathic score based on their performance.

After completing the intervention or control, students wrote qualitative reflections about their experience. Moreover, 1 to 2 weeks later, all students completed an objective structured clinical examination (OSCE). As previously reported, the primary outcome of the trial showed that MPathic-VR has a positive, statistically significant effect on medical student's proficiency in disclosing a new cancer diagnosis (intercultural

scenario) and practicing conflict resolution in an interprofessional scenario with an oncology nurse [31].

As the trial demonstrated effectiveness of MPathic-VR intervention, we turned our focus to understanding the mechanisms and differential experiences with the MPathic-VR intervention. The aim of this study was to focus on only students exposed to the MPathic-VR simulation in the intervention arm and investigate differential effects of the simulation. Our primary aim was to answer the question: *How do medical student reflections about their experiences compare between low, medium, and high performers on the primary outcome measures of communication performance in the simulation and on the OSCE?*

Methods

Design

This study employed a convergent mixed methods design involving integrating 2 quantitative measures (MPathic-VR–calculated scores and the OSCE scores) with qualitative reflections by learners about their experiences. Examining qualitative reflections helped to elucidate educational mechanisms for high, medium, and low OSCE scoring participants. The full randomized controlled trial results are reported elsewhere [31]. The University of Michigan Institutional Review Board (HUM00134766) determined that the study was exempt under the educational research category and waived documentation of consent for students.

Setting

The intervention took place at Eastern Virginia University Medical School, the University of Michigan Medical School, and the University of Virginia School of Medicine.

Participants

Second-year medical students from the 3 medical schools were recruited and randomly assigned to the control (n=211) or MPathic condition (n=210), the MPathic group is the focus of this study. We excluded 4 individuals because of missing qualitative data, leaving 206 students for this analysis.

Qualitative Data Collection

We employed an ethnographically driven approach for the qualitative component of the study as we sought to understand the nature of their experiences with the system. The students in each condition completed a reflective essay on their experience. Students were randomized to 1 of the 5 reflective questions about (1) human interactions, (2) understanding nonverbal communication, (3) most important things learned, (4) how to improve the simulation, and (5) functional aspects.

Qualitative Data Analysis

We imported all reflective responses into MAXQDA qualitative software (VERBI) to manage data for further analysis. We then followed a thematic qualitative text analysis [32] process. The first step was to review the entire database while memoing potential themes and connections among the data [33]. Next, we began coding the responses by assigning descriptive labels to segments of text and generated an initial codebook. A total of 3 individuals coded (RS, SB, and TG) data and met to review codes and reach agreement through a consensus process [34]. We paused to review the codes, discussed discrepancies to create a common code definition, eliminated redundant codes, and refined descriptions for each code. Using this codebook, we proceeded with systematically coding the remainder of the database but also allowed additional codes to emerge as needed. Finally, we grouped similar codes into themes and considered the relationship among themes.

Quantitative Data Collection

MPathic-VR scores (continuous data) were collected that reflect the path through the system and responses for each participant. For each exchange with the VH, the system recorded a point value (3-point scale of optimal to worst with lower scores reflecting better performance) and summed points for an overall MPathic-VR score for each scenario.

Students' advanced communication skills were assessed through an OSCE around a novel scenario, although all students knew they would be tested. Standardized patient instructors evaluated each student's performance on a 5-point scale across 4 domains: open or defensive, collaborative or competitive, nonverbal communication, and an awareness of others.

Quantitative Data Analysis

One score was missing for the interprofessional scenario. Using an analysis of variance, we compared OSCE scores for the MPathic-VR and control condition. In addition, we examined changes in the MPathic score upon first run through the scenario and the repeat run after receiving feedback. Statistical significance was determined by a 95% CI.

Mixed Methods Integration

The value of mixed methods research lies in meaningful integration of qualitative and quantitative components. We employed merging integration, which consisted of comparing qualitative findings with respect to OSCE and MPathic scores and determining whether the qualitative findings confirmed, disconfirmed, or expanded our understanding of the scores [35]. Data were not normally distributed, so we converted each OSCE and MPathic score into a 3-level categorical variable (high, medium, and low). On the basis of tertiles, OSCE scores of less than 0.57 were categorized as low, 0.57 to 0.95 as medium, and greater than 0.95 as high. MPathic scores from the second run-through for each scenario were converted into categorical variables. Keeping in mind that low MPathic scores reflect better performance, for the intercultural scenario, scores of less than 3 were categorized as high, 3 to 7 as medium, and greater than 7 as low. In the interprofessional scenario, scores of less than 3 were categorized as high, 3 to 6 as medium, and greater than 6 as low. We used joint display analysis by arraying qualitative themes by OSCE and MPathic scores to understand how the experience differed among participants at these OSCE and MPathic score levels.

Results

Qualitative Results

Our analysis identified 3 major positive themes for the MPathic-VR group learners: gaining useful communication skills, learning awareness of nonverbal skills in addition to verbal skills, and feeling motivated to learn more about communication. A subset of participants expressed some reservations about this initial version of the system encapsulated by subthemes of potential improvements: uncertainty about timing/suitability of training with the VHs in the system versus actual humans, questions about the *repetition*, and a minority disinterested in any communication training.

Gaining Useful Communication Skills

Regarding verbal communication, learners reported gaining strategies for effectively interacting with patients and other health care providers in real clinical settings. Frequently mentioned strategies included asking open-ended questions, validating or acknowledging the conversation partner's feelings with reflective language, and remembering the importance of a simple apology. A few reported that the time built in to consider their verbal response and the immediate feedback helped them carefully think about their word choices before speaking. Specifically, learners reported being more mindful of selecting words that display cultural humility (avoiding assumptions), using inclusive language (*we* or *together* vs *I*), and avoiding *yes but* sentences (nonconfrontational phrasing).

Learning Awareness of Nonverbal Skills

Learners reported becoming more cognizant of their facial expressions in conversation, especially eyebrow movement, nodding, and smiling. Some learned about their own unfavorable behavior that they were not previously aware of, such as twitching and fidgeting. Beyond personal awareness, learners mentioned that they became aware of the importance of nonverbal communication in establishing rapport and conveying interest, acknowledgment, and empathy.

Feeling Motivated to Learn More About Communication

Learners noted that they would benefit from more training like MPathic-VR and expressed interest in interacting with the system through different scenarios. Learners were interested in additional scenarios that involve noncompliant or angry patients, determining a care plan, and delivering difficult diagnoses.

Optimizing the System

A subset of participants expressed either some reservations about the initial version of the system or disinterest in communication training. As the themes represented a small proportion, we labeled them as subthemes. One subtheme is the uncertainty about VH rather than humans for training. A few learners commented on the mechanics of talking with a VH owing to the lack of variability in conversation because of the multiple-choice responses and the interruption in conversation and eye contact to read the prompts on screen. Another subtheme was that the training was *too repetitive* by having individuals practice talking and repeat each scenario after receiving feedback. These individuals reported losing interest in the exercise when having to complete the scenario for the second time.

Disinterest in the training included both feeling "I already know how to communicate" and doubting the importance of nonverbal behaviors. A few learners felt as if the lessons provided were "common sense" and did not expand on their current knowledge, reflecting a broader lack of interest in communication skill building. Some learners mentioned that the system's prompt to use nonverbal behavior made the interaction "phony."

Quantitative Results

The MPathic score for the group randomly assigned to the training intervention is based on responses through each scenario, and it indicated a statistically significant improvement from the initial to the repeat scenario after feedback. Scores improved (a lower score is better) from the first (mean 11.67, SD 6.26) to the second time through the intercultural scenario (mean 5.89, SD 5.12; $F_{1,207}$=166.14; P<.001), and scores improved for the interprofessional scenario from the first (mean 7.59, SD 3.96) to the second time (mean 4.62, SD 2.54; $F_{1,207}$=104.64; P<.001). The global OSCE score was better for the MPathic-VR condition than the control condition, as reported in detail previously ($F_{1,414}$=6.09; P=.01) [31]. The nonverbal subdomain was also significantly higher for the MPathic-VR condition ($F_{1,414}$=13.70; P<.001) [31].

Integrated Mixed Methods Results

Primary Aim: Objective Structured Clinical Examination Score and Qualitative Comments

The distribution of medical students according to lowest, medium, and highest scoring participants on the OSCE was 10.7% (22/206), 66.0% (136/206), and 23.3% (48/206), respectively. We examined whether any patterns were present among qualitative themes by the lowest, medium, and highest scoring participants on the OSCE, as reported in Table 1. Although we did not find differences for all themes, several noteworthy patterns emerged.

Learners in all 3 groups commented on learning useful verbal and nonverbal communication skills, although low OSCE scorers had fewer comments compared with medium and high OSCE scorers. Most of the variation between groups was noted in comments about nonverbal behaviors. Interestingly, several medium and high OSCE scorers specifically commented on learning how to use nonverbal communication where appropriate, whereas only 1 low scorer mentioned nonverbal aspects. Examples included using body language to build rapport and replacing verbal responses with nonverbal cues to avoid interrupting the flow of the conversation. Many leaners also commented on being mindful when nodding and smiling, noting that these behaviors can be misinterpreted in certain situations as insensitive or arrogant.

Table 1. A joint display of qualitative themes by quantitative performance level on an objective structured clinical examination.

Themes	Objective structured clinical examination advanced communication assessment		
	Low (<0.55)	Medium (0.54-0.98)	High (>0.98)
Useful communication skills	N/A[a]	"Effective communication both verbal and nonverbal will be essential in getting the best care for patients."	"I thought that I was given helpful strategies for interacting with patients such as asking open-ended questions, validating feelings, and types of nonverbal cues to use."
Remembering nonverbal skills	"Smiling and nodding is also important"	"Non-verbal cues can be very helpful. There are good times to nod and also times when it is not appropriate." and "In emotionally charged situations, I realize that using non-verbal communication is very important."	"Helped teach how to read facial expressions from people such as when the nurse was upset."
Motivated to learn more	N/A	"I would definitely benefit from more training such as this. I found myself hoping that there would be another simulation or two."	"It would be interesting to go through other scenarios, and to see if this actually has a positive effect on my future interactions with patients."
Prefer humans	"Hard to engage in non-verbal communication when you know you are just talking at a computer."	"I think that training for communication with patients is better done with live patients."	"Your true response can only come from human to human interaction…program is much stronger at allowing a person to think about their verbal responses."

[a]N/A: not applicable.

Medium and high OSCE scorers also reported gaining nonverbal skills in managing emotionally charged and complex situations (eg, family politics in health care), which was not discussed by low scorers. High scorers also noted the importance of mirroring their conversation partner's facial expressions to help diffuse tense situations. Low scorers did not provide comments that indicated awareness of their conversation partner's nonverbal behaviors.

Expressions of motivation to learn about communication were made only by medium and high OSCE scorers. However, reservations about training arose from some of these learners as well. Notably, several medium OSCE scorers and 1 high scorer questioned the value of the training relative to their time. They explained that they were unable to fully immerse themselves in the exercise because they were distracted by external factors such as exams. Comments about already knowing how to communicate were also only made by medium and high OSCE scorers.

Comments from some members of all 3 groups questioned the role of VHs for communication training versus humans. We found no difference among the 3 performance groups in discussing realism of the VH training experience.

Secondary Aim: MPathic Scores and Qualitative Comments

The distribution of medical students according to lowest, medium, and high scores on the intercultural MPathic simulation participants was 18.4% (38/206), 60.2% (124/206), and 21.4% (44/206), respectively. The distribution of medical students according to lowest, medium, and high scores on the interprofessional MPathic simulation participants was 18.9%) (39/206), 64.6% (133/206), and 16.0% (33/206), respectively. One score was missing for the interprofessional scenario. We investigated whether any patterns were present among qualitative themes by low, medium, and high scoring participants in MPathic scores for both intercultural and interprofessional scenarios. In both scenarios, patterns were similar to those that emerged when comparing low, medium, and high OSCE scorers against their qualitative experiences. High, medium, and low group membership overlapped for both the OSCE and MPathic scores (eg, high OSCE and MPathic scores). In the intercultural scenario, 12 learners were in the high scoring group for both OSCE and MPathic scores, 81 were in the middle scoring group, and 6 were in the low scoring group. In the interprofessional scenario, 6 were in the high scoring group, 87 were in the middle scoring group, and 4 were in the low scoring group for both OSCE and MPathic scores.

In both scenarios, learners across all groups acknowledged the use of appropriate nonverbal behaviors. Learners in all groups also mentioned that they learned how to use nonverbal behavior to help manage tense situations. However, when comparing low, medium, and high OSCE scores against qualitative themes, these were mentioned mostly by high OSCE performers and only 1 low OSCE performer.

When comparing the qualitative comments made by low, medium, and high scorers between the 2 MPathic scenarios, similar patterns emerged. For the intercultural scenario, comments indicating a desire to learn more about communication were made across all groups, although there were more mentions from high performers compared with low performers. In contrast, in the interprofessional scenario, low and middle performers had more of a desire to learn about communication compared with high performers.

Improvement in MPathic Scores and Qualitative Comments

Finally, we compared comments by whether learners improved their MPathic scores. The distribution of pre-post change for the intercultural scenario included improved scores (85.4%, 176/206), no change (7.3% (15/206), and worse performance

(7.3%, 15/206). For the interprofessional simulation, the distribution was 77.7% (160/206), 10.2% (21/206), and 11.7% (24/206), respectively. In both scenarios, learners who did not improve or those who did worse on the second run-through made comments about engagement in the training that related to a lack of interest. These learners indicated that they would rather use the MPathic training time to study for other courses. In contrast, those who improved their scores were more likely to mention engagement issues that related to training procedures, such as being unaccustomed to gesturing to a computer screen. In both scenarios, learners who did worse on the second run-through did not want to learn more or practice more.

Discussion

Principal Findings

Several qualitative differences in themes were apparent when comparing high, middle, and lower performing individuals based on posttraining OSCE communications performance scores. The higher scoring individuals noted the importance of learning about communication and communicating appropriately more than lower scoring individuals. Compared with lower performers, higher performers emphasized the importance of verbal and nonverbal communication skills when interacting in health settings. The pattern was especially notable in their reflection of nonverbal communication. Overall, the integrated results suggest that higher performing individuals seemed to understand and perhaps have stronger buy-in as to the importance of health communication and motivation to learn further.

Regarding our secondary aims of examining how learners compared based on their overall performance during the simulation, and on the first and second runs, learners who did not improve or performed worse on the second run-through in both scenarios expressed more dissatisfaction with the learning experience. These learners showed no interest in wanting to learn more about communication or practicing with the MPathic training system or perhaps did not understand the importance of communication training. This pattern was not as blatant when comparing qualitative themes by low, medium, and high performers in OSCE and MPathic scores. These learners may simply not be good candidates for the training because of personal factors, such as competing priorities. However, in an actual implementation, MPathic-VR could be available on demand at any time needed to fit the learner's schedule. In addition, these individuals may have had low motivation to participate, which had a negative effect on learning outcomes [36], for example, a subset doubted the importance of nonverbal skills. A potential strategy is to increase upfront information and explain that nonverbal expression transmits a larger portion of information than words. Ironically, such students may benefit the most from communications training.

Comparison With Prior Work

Recent research has examined the value of virtual patients, broadly, applied to improve medical decision making, such as creating accurate virtual patient cases derived from electronic health records to train decision making [37]. Our work further adds to the literature about VHs, focusing specifically on communication training, critical to decision making and the entire health care encounter. Our overall posttraining measures indicated a favorable benefit of VH training, similar to results of other VH interventions to improve verbal communication skills, such as VHs to enhance providers' ability to impart knowledge about appropriate antibiotic use and improve views of shared decision making [15]. Our study, however, also investigated effects on nonverbal communication skills and found promising outcomes. Although nonverbal skills were assessed, learner feedback is not yet automated, as in Liu et al's Web-based system used to provide automated nonverbal feedback [38].

The unique contributions of this study arise from using a mixed methods approach and focusing our attention on only students exposed to the MPathic-VR simulation to understand the mechanisms by which learners interacted with the intervention. In our investigation, merging of qualitative and quantitative databases revealed confirmation of findings related to intervention differential effects and mechanisms of action. Furthermore, visualization of these findings through the use of a joint display facilitated our interpretation of findings (Table 1). One mechanism related to the intervention was motivation. We found less favorable OSCE and MPathic scores among learners who expressed a lack of interest in the training in their written reflection. Thus, motivational issues seemed to be an important human factor related to engagement with the informatics technology. This result has implications in either priming health professionals before participating or selecting who is most likely to benefit. Another mechanism is awareness of nonverbal communication. Learners who explained that they gained skills in the appropriate use of nonverbal communication to manage emotionally charged situations tended to achieve better intervention outcomes (ie, favorable OSCE). Confirmation of findings with both quantitative and qualitative methods lends credibility to our understanding of these MPathic intervention mechanisms [35]. Furthermore, results suggest that humans will engage in nonverbal communication when interacting with an informatics-driven intelligent VH. When appropriate, nonverbal communication might be considered among the human factors principles in conducting research on interaction with informatics technology.

Study Limitations

Our study was limited to medical students, and future research is needed to apply the simulation to practicing providers and other health professionals. In addition, nonverbal communication was not the focus of data collection despite arising as a primary finding. Training nonverbal communication skills in VH simulation merits further investigation. In addition, the nonverbal behaviors specifically selected for this simulation—nodding, smiling, and eyebrow raises—are a few of the many possible nonverbal and verbal behaviors associated with empathy expression [39]. Coding schemes that offer methods for meticulously analyzing nonverbal behaviors and their associated emotions exist and can be applied to the Kinect sensor data collected in the trial [40,41].

Conclusions

Future research should address how to motivate and engage learners. On one hand, the MPathic system has made an innovative leap forward by demonstrating the effect of the system to train communication skills that transfer into a realistic communication scenario. The results raise research questions about the need to incorporate instructional design principles that will help motivate students skeptical about improving their health care communication [42]. Although the results indicated that higher performers had stronger beliefs about the importance of good communication in health care, it is unclear whether these differences were present before participating in the intervention. Future research may benefit from additional pretest measures, such as a written reflection about communication and empathy assessments, such as the Jefferson Scale [43].

Acknowledgments

This study was supported by the National Institutes of Health (1K01LM012739).

Conflicts of Interest

FK serves as president of Medical Cyberwords, which develops MPathic-VR, and MF has stock options in Medical Cyberworlds. The University of Michigan Conflict of Interest Office considered potential for conflict of interest and concluded that no formal management plan was required.

References

1. Beckman HB, Markakis KM, Suchman AL, Frankel RM. The doctor-patient relationship and malpractice. Lessons from plaintiff depositions. Arch Intern Med 1994 Jun 27;154(12):1365-1370. [doi: 10.1001/archinte.154.12.1365] [Medline: 8002688]
2. Curry LA, Spatz E, Cherlin E, Thompson JW, Berg D, Ting HH, et al. What distinguishes top-performing hospitals in acute myocardial infarction mortality rates? A qualitative study. Ann Intern Med 2011 Mar 15;154(6):384-390 [FREE Full text] [doi: 10.7326/0003-4819-154-6-201103150-00003] [Medline: 21403074]
3. Epstein RM, Franks P, Shields CG, Meldrum SC, Miller KN, Campbell TL, et al. Patient-centered communication and diagnostic testing. Ann Fam Med 2005;3(5):415-421 [FREE Full text] [doi: 10.1370/afm.348] [Medline: 16189057]
4. Levinson W, Roter DL, Mullooly JP, Dull VT, Frankel RM. Physician-patient communication. The relationship with malpractice claims among primary care physicians and surgeons. J Am Med Assoc 1997 Feb 19;277(7):553-559. [doi: 10.1001/jama.277.7.553] [Medline: 9032162]
5. Ong LM, de Haes JC, Hoos AM, Lammes FB. Doctor-patient communication: a review of the literature. Soc Sci Med 1995 Apr;40(7):903-918. [doi: 10.1016/0277-9536(94)00155-m] [Medline: 7792630]
6. Andel C, Davidow SL, Hollander M, Moreno DA. The economics of health care quality and medical errors. J Health Care Finance 2012;39(1):39-50. [Medline: 23155743]
7. Holzinger A, Langs G, Denk H, Zatloukal K, Müller H. Causability and explainabilty of artificial intelligence in medicine. WIREs Data Mining Knowl Discov 2019:e1312. [doi: 10.1002/widm.1312]
8. Davis MH. Empathy: A Social Psychological Approach. New York: Routledge; 2018.
9. Lim BT, Moriarty H, Huthwaite M. 'Being-in-role': A teaching innovation to enhance empathic communication skills in medical students. Med Teach 2011;33(12):e663-e669. [doi: 10.3109/0142159X.2011.611193] [Medline: 22225448]
10. Spiro H. What is empathy and can it be taught? Ann Intern Med 1992 May 15;116(10):843-846. [doi: 10.7326/0003-4819-116-10-843] [Medline: 1482433]
11. Hojat M, Vergare MJ, Maxwell K, Brainard G, Herrine SK, Isenberg GA, et al. The devil is in the third year: a longitudinal study of erosion of empathy in medical school. Acad Med 2009 Sep;84(9):1182-1191. [doi: 10.1097/ACM.0b013e3181b17e55] [Medline: 19707055]
12. Griffin SJ, Kinmonth A, Veltman MW, Gillard S, Grant J, Stewart M. Effect on health-related outcomes of interventions to alter the interaction between patients and practitioners: a systematic review of trials. Ann Fam Med 2004;2(6):595-608 [FREE Full text] [doi: 10.1370/afm.142] [Medline: 15576546]
13. Fernndez-Olano C, Montoya-Fernndez J, Salinas-Snchez AS. Impact of clinical interview training on the empathy level of medical students and medical residents. Med Teach 2008;30(3):322-324. [doi: 10.1080/01421590701802299] [Medline: 18509879]
14. Patel V, Kannampallil T. Cognitive informatics in biomedicine and healthcare. J Biomed Inform 2015 Feb;53:3-14 [FREE Full text] [doi: 10.1016/j.jbi.2014.12.007] [Medline: 25541081]
15. Schoenthaler A, Albright G, Hibbard J, Goldman R. Simulated conversations with virtual humans to improve patient-provider communication and reduce unnecessary prescriptions for antibiotics: a repeated measure pilot study. JMIR Med Educ 2017 Apr 19;3(1):e7 [FREE Full text] [doi: 10.2196/mededu.6305] [Medline: 28428160]
16. Mantovani F, Castelnuovo G, Gaggioli A, Riva G. Virtual reality training for health-care professionals. Cyberpsychol Behav 2003 Aug;6(4):389-395. [doi: 10.1089/109493103322278772] [Medline: 14511451]

17. Bickmore T, Marsella S, Sidner C. Intelligent Virtual Agents. In: Proceedings of the 14th International Conference. 2014 Presented at: IVA'14; August 27-29, 2014; Boston, MA, USA p. 14. [doi: 10.1007/978-3-319-09767-1]
18. Kron F. Medica ex Machina: Can You Really Teach Medical Professionalism with a Machine? Journal of Clinical Oncology/ASCO Educational Supplement Spring 2007:692-697. [doi: 10.2307/j.ctt20q245q.6]
19. Kron FW, Gjerde CL, Sen A, Fetters MD. Medical student attitudes toward video games and related new media technologies in medical education. BMC Med Educ 2010 Jun 24;10:50 [FREE Full text] [doi: 10.1186/1472-6920-10-50] [Medline: 20576125]
20. Lynch-Sauer J, Vandenbosch TM, Kron F, Gjerde CL, Arato N, Sen A, et al. Nursing students' attitudes toward video games and related new media technologies. J Nurs Educ 2011 Sep;50(9):513-523. [doi: 10.3928/01484834-20110531-04] [Medline: 21627050]
21. Kleinsmith A, Rivera-Gutierrez D, Finney G, Cendan J, Lok B. Understanding empathy training with virtual patients. Comput Human Behav 2015 Nov 1;52:151-158 [FREE Full text] [doi: 10.1016/j.chb.2015.05.033] [Medline: 26166942]
22. Lane HC, Hays MJ, Core MG, Auerbach D. Learning intercultural communication skills with virtual humans: feedback and fidelity. J Educ Psychol 2013 Nov;105(4):1026-1035. [doi: 10.1037/a0031506]
23. Reeves B, Nass C. The Media Equation: How People Treat Computers, Television, And New Media Like Real People And Places. New York, NY: Cambridge; 1996.
24. Moreno R, Mayer RE. Personalized messages that promote science learning in virtual environments. J Educ Psychol 2004 Mar;96(1):165-173. [doi: 10.1037/0022-0663.96.1.165]
25. Gratch J, Wang N, Okhmatovskaia A, Lamothe F, Morales M, van der Werf RJ. Can Virtual Humans Be More Engaging Than Real Ones? In: Proceedings of the 12th international conference on Human-computer interaction: intelligent multimodal interaction environments. 2007 Presented at: HCI'07; July 22-27, 2007; Beijing, China p. 286-297. [doi: 10.1007/978-3-540-73110-8_30]
26. Shute VJ. Focus on formative feedback. Rev Edu Res 2008;78(1):153-189. [doi: 10.3102/0034654307313795]
27. Kolb D. Experiential Learning: Experience as the Source of Learning and Development. Upper Saddle River, NY: Pearson; 1984.
28. Kolb AY, Kolb DA. The learning way: meta-cognitive aspects of experiential learning. Simul Gaming 2009;40(3):297-327. [doi: 10.1177/1046878108325713]
29. Ericsson KA. Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. Acad Med 2004 Oct;79(10 Suppl):S70-S81. [doi: 10.1097/00001888-200410001-00022] [Medline: 15383395]
30. Fetters M, Guetterman T, Scerbo M, Kron F. A two-phase mixed methods project illustrating development of a virtual human intervention to teach advanced communication skills and a subsequent blinded mixed methods trial to test the intervention for effectiveness. Int J Mult Res Appr 2018;10(1):296-316 [FREE Full text] [doi: 10.29034/ijmra.v10n1a19]
31. Kron FW, Fetters MD, Scerbo MW, White CB, Lypson ML, Padilla MA, et al. Using a computer simulation for teaching communication skills: a blinded multisite mixed methods randomized controlled trial. Patient Educ Couns 2017 Apr;100(4):748-759 [FREE Full text] [doi: 10.1016/j.pec.2016.10.024] [Medline: 27939846]
32. Kuckartz U. Qualitative Text Analysis: A Guide to Methods, Practice and Using Software. London: Sage; 2014.
33. Creswell JW, Guetterman TC. Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research. Sixth Edition. Upper Saddle River, NJ: Pearson; 2019.
34. Creswell JW. 30 Essential Skills for the Qualitative Researcher. Thousand Oaks, CA: Sage; 2015.
35. Fetters MD, Curry LA, Creswell JW. Achieving integration in mixed methods designs-principles and practices. Health Serv Res 2013 Dec;48(6 Pt 2):2134-2156 [FREE Full text] [doi: 10.1111/1475-6773.12117] [Medline: 24279835]
36. Shell D, Brooks D, Trainin G, Wilson K, Kauffman D, Herr L. The Unified Learning Model: How Motivational, Cognitive, And Neurobiological Sciences Inform Best Teaching Practices. Dordrecht: Springer; 2009.
37. Bloice MD, Simonic K, Holzinger A. Casebook: a virtual patient iPad application for teaching decision-making through the use of electronic health records. BMC Med Inform Decis Mak 2014 Aug 7;14:66 [FREE Full text] [doi: 10.1186/1472-6947-14-66] [Medline: 25100051]
38. Liu C, Lim RL, McCabe KL, Taylor S, Calvo RA. A web-based telehealth training platform incorporating automated nonverbal behavior feedback for teaching communication skills to medical students: a randomized crossover study. J Med Internet Res 2016 Sep 12;18(9):e246 [FREE Full text] [doi: 10.2196/jmir.6299] [Medline: 27619564]
39. Henry SG, Fuhrel-Forbis A, Rogers MA, Eggly S. Association between nonverbal communication during clinical interactions and outcomes: a systematic review and meta-analysis. Patient Educ Couns 2012 Mar;86(3):297-315. [doi: 10.1016/j.pec.2011.07.006] [Medline: 21824738]
40. Ekman P, Friesen W, Hager J. Facial Action Coding System: The Manual. Salt Lake City, UT: Research Nexus; 2002.
41. Ekman P, Rosenberg EL. What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS). New York: Oxford University Press, USA; 1997.
42. Cook DA, Artino AR. Motivation to learn: an overview of contemporary theories. Med Educ 2016 Oct;50(10):997-1014 [FREE Full text] [doi: 10.1111/medu.13074] [Medline: 27628718]

43. Hojat M, Mangione S, Nasca T, Cohen M, Gonnella J, Erdmann J, et al. The jefferson scale of physician empathy: development and preliminary psychometric data. Edu Psychol Measure 2001;61(2):349-365. [doi: 10.1177/00131640121971158]

Abbreviations

OSCE: objective structured clinical examination
VH: virtual human

Edited by G Eysenbach; submitted 19.07.19; peer-reviewed by S Erdley, A Holzinger; comments to author 15.09.19; revised version received 27.09.19; accepted 28.09.19; published 27.11.19.

Please cite as:
Guetterman TC, Sakakibara R, Baireddy S, Kron FW, Scerbo MW, Cleary JF, Fetters MD
Medical Students' Experiences and Outcomes Using a Virtual Human Simulation to Improve Communication Skills: Mixed Methods Study
J Med Internet Res 2019;21(11):e15459
URL: http://www.jmir.org/2019/11/e15459/
doi:10.2196/15459
PMID:31774400

Original Paper

Development and Evaluation of a New Serious Game for Continuing Medical Education of General Practitioners (Hygie): Double-Blinded Randomized Controlled Trial

Louis-Baptiste Jaunay[1,2], MSc, MD; Philippe Zerr[2], MD; Lino Peguin[2]; Léandre Renouard[2]; Anne-Sophie Ivanoff[2], MD; Hervé Picard[3,4], MD; James Griffith[5], PhD; Olivier Chassany[6,7], MD, PhD; Martin Duracinsky[6,7,8], MD, PhD

[1]Département de Médecine Générale, Sorbonne Paris Cité, Université Paris-Descartes, Paris, France
[2]Département de Médecine Générale, Sorbonne Paris Cité, Université Paris-Diderot, Paris, France
[3]Service de Recherche Clinique, Fondation Rothschild, Paris, France
[4]Pole Recherche et Evaluation Scientifique, Cabinet Ipso, Paris, France
[5]Department of Medical Social Sciences, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States
[6]Patient-Reported Outcomes Research, Sorbonne Paris Cité, Université Paris-Diderot, Paris, France
[7]Unité de Recherche Clinique en Economie de la Santé, Hôpital Hôtel-Dieu, Assistance Publique Hôpital de Paris, Paris, France
[8]Médecine Interne et Immunologie Clinique, Hôpital Bicêtre, Assistance Publique Hôpital de Paris, Paris, France

Corresponding Author:
Louis-Baptiste Jaunay, MSc, MD
Département de Médecine Générale
Sorbonne Paris Cité
Université Paris-Descartes
24 Rue du Faubourg Saint-Jacques
Paris, 75014
France
Phone: 33 144412363
Fax: 33 144412364
Email: drjaunay@gmail.com

Abstract

Background: Continuing medical education is important but time-consuming for general practitioners (GPs). Current learning approaches are limited and lack the ability to engage some practitioners. Serious games are new learning approaches that use video games as engaging teaching material. They have significant advantages in terms of efficiency and dissemination.

Objective: The aim of this study was to create a serious game and to evaluate it in terms of effectiveness and satisfaction, comparing it with a traditional method of continuing education—article reading.

Methods: We produced a prototype video game called *Hygie* on the 5 most common reasons of consultation in general practice using 9 articles from independent evidence-based medicine journals (reviews from *Prescrire* and *Minerva*). We created 51 clinical cases. We then conducted a double-blinded randomized trial comparing the learning provided by a week of access to the game versus source articles. Participants were GPs involved as resident supervisors in 14 French university departments of family practice, recruited by email. Primary outcomes were (1) mean final knowledge score completed 3 to 5 weeks after the end of the intervention and (2) mean difference between knowledge pretest (before intervention) and posttest (3 to 5 weeks after intervention) scores, both scaled on 10 points. Secondary outcomes were transfer of knowledge learned to practice, satisfaction, and time spent playing.

Results: A total of 269 GPs agreed to participate in the study. Characteristics of participants were similar between learning groups. There was no difference between groups on the mean score of the final knowledge test, with scores of 4.9 (95% CI 4.6-5.2) in the *Hygie* group and 4.6 (95% CI 4.2-4.9) in the reading group (P=.21). There was a mean difference score between knowledge pre- and posttests, with significantly superior performance for *Hygie* (mean gain of 1.6 in the *Hygie* group and 0.9 in the reading group; P=.02), demonstrating a more efficient and persistent learning with Hygie. The rate of participants that reported to have used the knowledge they learned through the teaching material was significantly superior in the *Hygie* group: 77% (47/61) in the *Hygie* group and 53% (25/47) in the reading group; odds ratio 2.9, 95% CI 1.2-7.4. Moreover, 87% of the opinions were favorable,

indicating that *Hygie* is of interest for updating medical knowledge. Qualitative data showed that learners enjoyed *Hygie* especially for its playful, interactive, and stimulating aspects.

Conclusions: We conclude that *Hygie* can diversify the offering for continuing education for GPs in an effective, pleasant, and evidence-based way.

Trial Registration: ClinicalTrials.gov NCT03486275; https://clinicaltrials.gov/ct2/show/NCT03486275

(J Med Internet Res 2019;21(11):e12669) doi:10.2196/12669

KEYWORDS

general practice; continuing medical education; evidence-based medicine; video games; randomized controlled trial; pedagogy

Introduction

Background

General practitioners (GPs) update their medical knowledge throughout their professional life to maintain knowledge acquired during their initial studies and to be abreast of the latest scientific advances.

Continuing medical education, however, can be tedious and sporadic because a considerable amount of new medical data and new literature are being continuously released, varying in quality and accessibility. The busy practitioner has limited time to consult this information [1-3], and traditional teaching methods such as lectures and group discussion have small and short-lasting effects [4]. As a result, clinical care may not be in line with the latest science, leading to poorer health outcomes [5]. Thus, new, efficient, and stimulating teaching methods are required.

New teaching materials called *serious games* are efficient [6,7] and easily disseminated methods for education [8]. Indeed, they offer the possibility of combining learning activities such as testing [9], feedback [10], spaced repetition [11], and problem-based learning [12,13] with a positive experience. Learning challenges can be provided by these games [14,15] in a risk-free environment [16]. Therefore, serious games give active participation and autonomy to the learner, both of which are crucial qualities in adult education [17].

Few serious games have been developed with the goal of facilitating continuous medical education for health professionals [18] and GPs [19,20]. To our knowledge, no existing game covers several topics related to family medicine.

Objectives

The aim of this study was to develop a prototype of a new serious game called *Hygie* for continuing medical education for the GP and to assess its effectiveness and user acceptance as compared with a traditional activity (article reading) in a randomized trial.

Methods

Design and Development of Hygie

We produced a prototype video game called *Hygie* in which the player is a GP in the process of treating several patients. We defined topics for this prototype based on the 5 most frequent reasons for consultation in France [21]: hypertension, health check and prevention, dyslipidemia, acute fever, and rhinopharyngitis.

For these 5 topics, we reviewed 9 articles in 2 French evidence-based journals: 6 from *Prescrire* and 3 from *Minerva* [22-30]. We selected these 2 journals because they provide robust evidence-based recommendations and are strictly independent from industrial and institutional influences.

From these 9 articles, we created 51 short clinical cases, each having 1 question that could be answered either by multiple choice or free text.

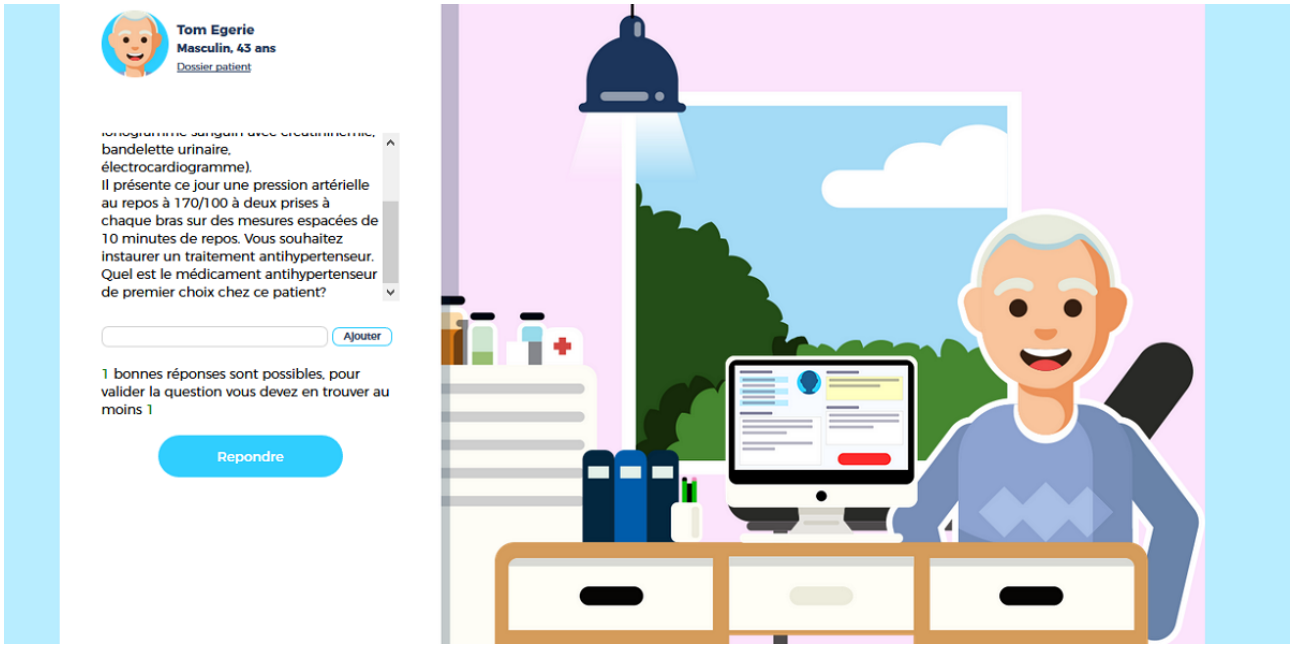
The game was coded using HTML 5, Cascading Style Sheets 3, JavaScript (ECMAScript 2015), and Hypertext Preprocessor (PHP) 7. Graphics were created using Adobe Illustrator and Adobe Photoshop (Figure 1).

Learning methods incorporated into the game included statement of educational objectives, immersion in a general medical consultation setting, problem-based learning with active restitution of knowledge, spaced recall, stimulation of intrinsic motivation by earning points, and having goals and levels with a “final boss” for each level. Humoristic elements such as puns in patients’ names were included to maximize engagement.

A preliminary test phase was conducted with 11 GPs and 9 residents in general practice. The preliminary test allowed us to detect and solve bugs, clarify questions, and sort questions into 5 levels of difficulty.

The prototype of the game is freely accessible on the Web [31].

Figure 1. Start of a clinical case in *Hygie*.



Study Design

We performed a double-blind randomized controlled trial to assess the effectiveness of *Hygie* as a method for continuing education for GPs as compared with a traditional article reading activity with the same content.

We asked all 35 French university departments of family practice to contact GPs involved as resident supervisors by email to participate in a real-life experience learning where they would have access to an electronic learning (e-learning) teaching material for 7 days, without mentioning the nature of the teaching materials. Institutional affiliations of the investigators were indicated at the end of the email.

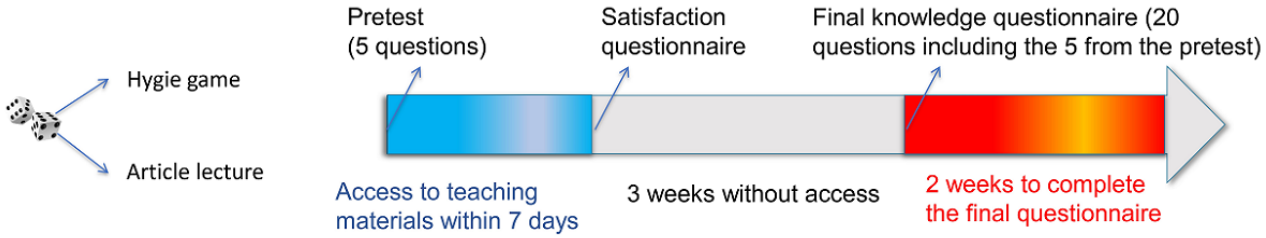
Information was delivered to participants about the purpose, the duration, the time to devote to the study, and anonymization of results.

After agreeing to participate, GPs accessed the study website where they completed a demographic questionnaire, a knowledge pretest of 5 questions on each of the 5 reasons for consultation. They were randomized using the rand function of PHP language (allocation ratio 1:1) to either the intervention group (Web access to *Hygie* for 1 week) or to the control group (access online to the 9 articles). Participants had an individual login, allowing them to access only the teaching material assigned to them. They did not know if they were assigned to the intervention or control group and did not know which intervention was performed in the other group. Data were collected on a Structured Query Language database.

After 1 week of free access to their respective teaching material (serious game *Hygie* vs articles), access was terminated. Reminders were sent to the 2 groups within 3 and 6 days of access to teaching material.

After 3 weeks without access to the teaching materials, participants received a final, 20-item knowledge questionnaire (Figure 2). Among the 20 questions, 5 were common with pretest. Only those participants who had completed the final questionnaire were analyzed.

Figure 2. Study design.



Primary outcomes were (1) dynamic and (2) static knowledge assessed by questionnaires:

1. A Dynamic Questionnaire-5 (DQ-5) with 5 items compared individual change of score between pretest (before intervention) and posttest (3-5 weeks after intervention). It was a 5-item questionnaire with each of the 5 questions weighted by a scale ranging from 1 to 3 according to its importance for practice and a global score from 0 to 14.

The goal of this questionnaire was to assess progression of each participants. For a simpler interpretation, we scaled the DQ-5 score to be out of 10 rather than 14.

2. A Static Questionnaire-20 (SQ-20) measured mean final score 3 to 5 weeks after intervention. It was a 20-item questionnaire (5 of the dynamic questionnaire plus 15 other questions), with each of the 20 questions weighted by a scale ranging from 1 to 3 according to its importance for practice and a global score from 0 to 58. The goal of this questionnaire was to compare groups, minimizing the potential carryover effects induced by the pretest questionnaire. Like for the DQ-5, for interpretation, we rescaled the SQ-20 to a 0 to 10 scale (rather than 0-58).

Here is an example of a knowledge question that appears in both dynamic and static questionnaires and the scoring method:

Question: Which cholesterol-lowering drugs have shown a decrease in mortality and morbidity?

Expected answers (free text): pravastatin, simvastatin.

Scoring method: It was rated 3 points out of 14;

If the 2 right molecules (pravastatin and simvastatin) are mentioned: 3 points

If 1 good molecule among pravastatin and simvastatin is mentioned: 1 point

In all other cases: 0 points.

The 2 knowledge questionnaires and their scale were written from the source articles by 3 experienced physicians who had no information about the game content, with instructions to identify practice-relevant issues in the articles. Participants' questionnaires were scored blindly by a physician not involved in the other stages of the study.

Secondary outcomes were (1) the use in medical practice of the knowledge acquired through the teaching material assessed at the time of the final questionnaire (participants answered the question "In the course of your practice, did you use the knowledge you learned through the teaching material?"), (2) time spent playing by participants assigned to *Hygie*, and (3) a satisfaction questionnaire. The satisfaction questionnaire, composed of 8 questions and completed at the end of the 1-week learning period, included quantitative and qualitative data about participant satisfaction, time reported as spent on the materials, and additional demographic data (eg, workplace and usual training materials for continuing education). Qualitative data were analyzed by content for themes related to effective learning as well as to illuminate potential strengths and weaknesses of *Hygie*. The average total time spent on the *Hygie* game was measured via server usage data. Average total time spent on the articles was not collected because participants could download the articles and read it offline.

Statistical Analysis

The answers to the knowledge and satisfaction questionnaires were collected on the framaform website. Statistical analyses were performed using R software (R Foundation for Statistical Computing) [32]. The 2 groups were compared using Fisher exact tests for nominal variables and Welch *t* tests for quantitative variables. Differences with $P<.05$ were considered significant.

Sample Size

The number of participants required with 80% power (1−beta) and 5% type I error was estimated before the study. A total of 128 participants were needed to detect a difference of 2 points out of 10 between the groups on the final questionnaire, assuming that the participants in the *Hygie* group had a final score of 8 out of 10 on average.

Ethics

Participation was anonymous and voluntary. Participants began the study by clicking a link to teaching materials.

The study was approved by the Committee for the Evaluation of the Ethics of Research Projects of hospital Robert Debré n° 2017/359.

Results

Participant Statistics

A total of 14 university departments from 8 French regions accepted to participate in this study. A total of 3398 GPs were invited to participate in this study by email. Of these, 269 participants (7.9%) accepted to participate in the study. Recruitment occurred between May 31, 2017, and June 27, 2017. A total of 108 participants completed the study and were analyzed. There was no difference of baseline characteristics between participants who completed and participants who did not complete the study.

The inclusion flow diagram according to Consolidated Standards of Reporting Trials recommendations [33] is shown in Figure 3.

Baseline characteristics of participants in both groups were comparable (Table 1). Mean age was 40.9 years, there was a majority of women, and an urban setting was the most common. The most widely used continuing education method was reading print journals.

The DQ-5 pretest mean score was identical in the 2 groups: 3.4 (95% CI 2.9-3.8) in the intervention group and 3.8 (95% CI 3.2-4.3) in the control group ($P=.27$, not significant).

Average time between stopping access to support and completing the final questionnaire was similar in the 2 groups: 25.3 days (95% CI 24.2-26.5) in the *Hygie* group and 27.5 days (95% CI 26.3-28.7) in the control group.

Figure 3. Flow diagram.

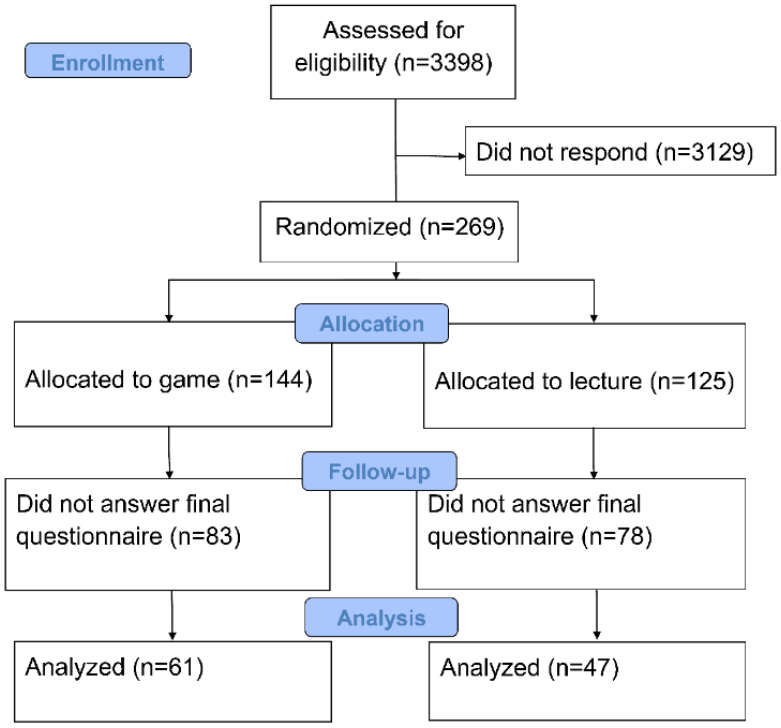


Table 1. Characteristics of analyzed participants allocated to each intervention (*Hygie* or control) at baseline (n=108).

Characteristics of participants	*Hygie* group (n=61)	Reading group (n=47)
Gender, n (%)		
Female	30 (49)	27 (57)
Male	31 (51)	20 (43)
Mean age (min-max)	39.8 (27-67)	42.4 (28-64)
Dynamic Questionnaire-5 pretest, mean score (95% CI)	3.4 (2.9-3.8)	3.8 (3.2-4.3)
Workplace setting, n (%)		
Rural	31 (51)	35 (74)
Semirural	23 (38)	7 (15)
Urban	7 (11)	6 (13)
Continuous teaching material, n (%)		
Paper journals	51 (84)	39 (83)
Internet journals	23 (38)	23 (49)
Internet sites	36 (59)	33 (70)
Onsite courses	48 (79)	33 (70)
Peer group training	35 (57)	20 (43)
Medical visitors	5	12

Outcomes

Knowledge

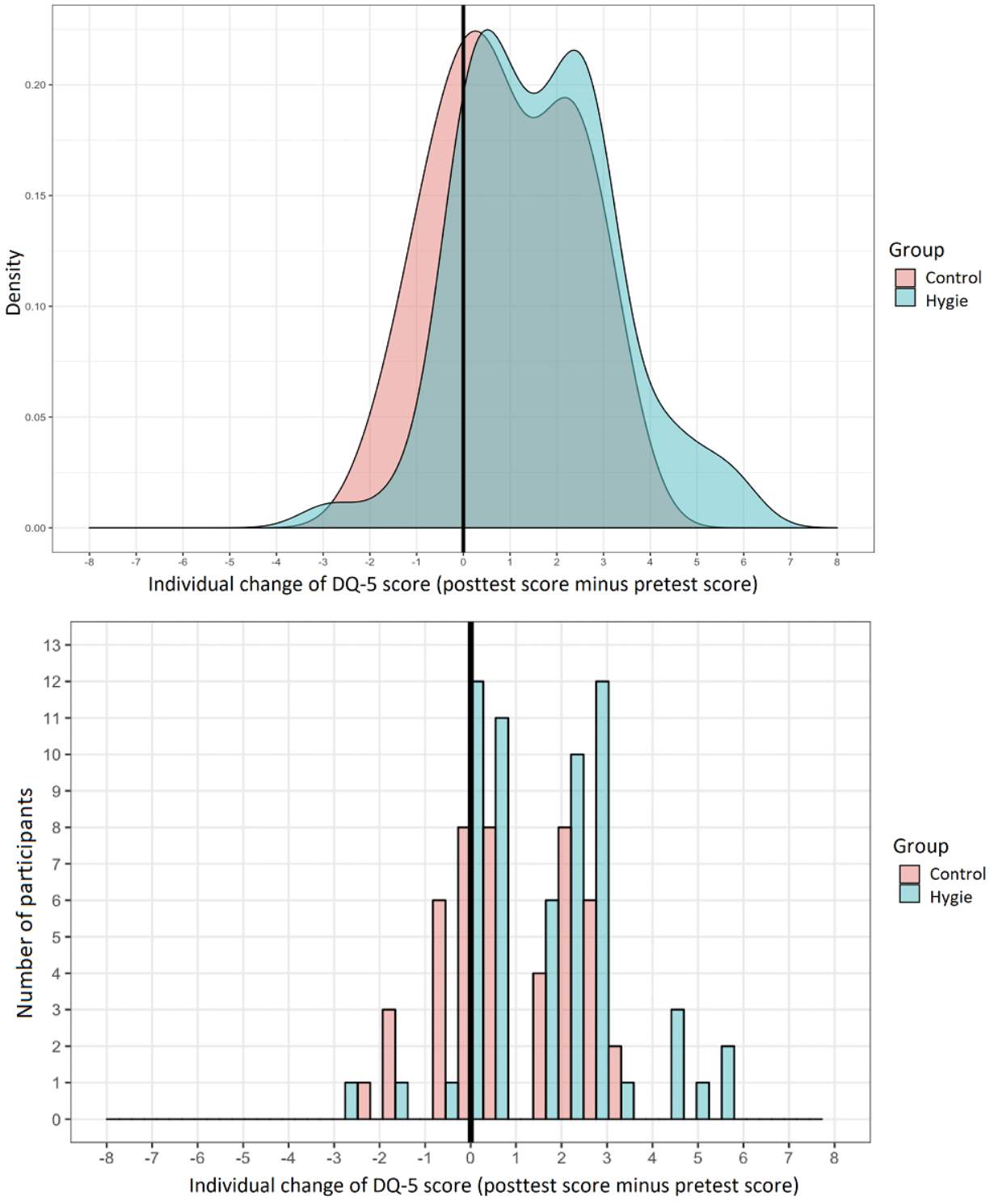
The final SQ-20 mean score was similar in the 2 groups: *Hygie* group 4.9 (95% CI 4.6-5.2) and control group 4.6 (95% CI 4.2-4.9; *P*=.21, not significant).

The final DQ-5 mean score (5-item posttest) was also similar in the 2 groups: *Hygie* group 5.0 (95% CI 4.6-5.4) and control group 4.7 (95% CI 4.2-5.1; *P*=.26, not significant).

The mean individual change of DQ-5 score between pre- and posttest was significantly superior to 0 in the *Hygie* group with a mean gain of 1.6 (95% CI 1.2-2.1; *P*<.001) and in control group with a mean gain of 0.9 (95% CI 0.5-1.4; *P*<.001).

For the critical test of our trial, this mean individual change of DQ-5 score between pre- and posttest at 3 to 5 weeks was significantly superior in the *Hygie* group compared with the reading group, with a difference of 0.7 (95% CI 0.1-1.3; *P*=.02; Figure 4).

Figure 4. Individual change of score in both groups, shown as density and number of participants. DQ-5: Dynamic Questionnaire-5.



Transfer to practice

For the question "In the course of your practice, did you use the knowledge that you learned through the teaching material?," the percentage of participants reporting "yes" was significantly greater in the *Hygie* group (77% in the *Hygie* group vs 53% in the reading group; odds ratio 2.9, 95% CI 1.2-7.4; Table 2).

Table 2. General practitioners' responses to "In the course of your practice, did you use the knowledge you learned through the teaching material?"

Response	*Hygie* group, n (%)	Reading group, n (%)
Yes	47 (77)	25 (53)
No	14 (23)	22 (47)

Players could rate clinical cases in terms of "usefulness to practice" just after resolving the cases. A total of 1464 clinical case scores were given, and the average score given was 4.14 out of 5.

Satisfaction and Qualitative Data

In the satisfaction survey, 87% of *Hygie* group participants answered "yes" to the question "Do you think that *Hygie* is of interest for updating your medical knowledge?" and 75% answered "yes" to the question "Do you think that *Hygie* should be allowed for continuing education credits?".

The qualitative reasons spontaneously mentioned by the participants also justified *Hygie*, including the following themes:

- Effective learning: the characteristics of the game (subthemes mentioned the following: speed, simple learning, and effective information assimilation), informative content (key messages, relevance of themes, clarity, and referenced responses), and its mechanisms (repetition of clinical cases promoting memorization, cognitive conflict that allows for better memorization, and allows one to learn test with real-life scenarios).
- An enjoyable experience (subthemes mentioned the following: playful and fun) with stimulating challenges (challenging stimulation and real-time style mimics the clinic): 36% of participants of the *Hygie* group answered "yes" to the question "Did this session make you want to consult medical journals more regularly or take out a subscription?," which suggests that gaming encourages players to read journals, considering that 73% of GPs already reported consulting *Prescrire* regularly and 12% reported consulting Minerva.

Time Spent on Supports

The average total time spent on the *Hygie* game, measured via server usage data for the included participants, was 43 min. The average time per game session was 10 min and 50 seconds. Participants self-reported the time they spent on learning materials in the satisfaction questionnaire through a discontinuous quantitative variable. The most common responses were "45 to 60 minutes" in the *Hygie* group and "10 to 20 minutes" in the reading group.

Success Rate and Comments on Game Questions

The overall success rate for clinical cases was 67%. Participants' comments on clinical cases reflected the cognitive conflict produced in players by the system of interaction between the GPs knowledge and the "model" proposed by the game: some agreed with the answer (eg, "bravo"), whereas others criticized the clarity of the question (eg, "one could specify [...]") or criticized the answer based on their practice (eg, "I would have liked to do [...] before treating") or other sources (eg, "recommendations on this topic include [...]").

Discussion

Principal Findings

To our knowledge, *Hygie* is the first continuing education material of this type; it is the first educational video game developed for and by GPs. *Hygie* was created without external funding and independently of the pharmaceutical and medical device industries. Moreover, it is based on reliable sources that are helpful to GPs in maintaining and expending their knowledge. Finally, it is unique because of its extensive evaluation among a significant number of GPs from several regions of France. The use of both a double-blinded randomized trial and a satisfaction questionnaire evaluation differentiates *Hygie* from other serious health games in existence, with a few exceptions such as InsuOnline [20]. Our study shows that it is feasible to create an engaging educational video game, including validation in a randomized trial, without influence of public or private financing.

Our results have shown that giving access to the *Hygie* game to GPs in "real life" conditions (ie, where learner decided when, where, and how much time he or she wants to spend learning) results in a persistent learning at 3 to 5 weeks. Furthermore, giving access to *Hygie* resulted in a better improvement in medical knowledge compared with giving access to articles, which is the traditional method. In addition, this knowledge seems to be more easily transferable to medical practice, as shown by the greater proportion of GPs reporting having used the knowledge in their own practices as compared with traditional journal article reading. This result suggests that serious games may engender better transfer of knowledge to real-life situations by actively engaging the learner.

No significant difference was found on the final questionnaire score, which is consistent with a previous study [20] and may suggest that journal article reading can still lead to sufficient knowledge for continuing education but that *Hygie* is at least noninferior to traditional methods.

Limitations and Strengths

There were some limitations to our study.

Recruitment was limited to GPs who were resident supervisors. This population is representative of the French GP population with some particularities such as a higher proportion of women, an underrepresented 45 to 54 years age group, a majority group practice, and a lower weekly working time [34]. Another bias is that participants were volunteered for the study after reading the email solicitation that offered to try a "new continuing education material." Thus, it was possible that this population of GPs was especially interested in updating their medical knowledge; this is supported by the proportion of physicians declaring reading the *Prescrire* journal in our study (70%), which is much higher than the proportion of French GPs

subscribing to *Prescrire* (18.1% of GPs subscribed to *Prescrire* in 2016) [35].

The GPs' positive response rate for participating in the study was 7.9% (269 included out of 3398 requested), which is comparable with the average response rate in this population [36] but prevented us from reaching the number of participants suggested by our power analysis. The real response rate cannot be definitively known because it is possible that some emails failed to reach potential GPs and were not read.

Contamination bias between groups is a potential limitation, but limited access to 1 of the 2 teaching materials through the login and individual working environment of French GPs has limited this possibility.

More than half (60%) of the participants did not complete the study, which may have been a consequence of the "real life" conditions of our trial (unconstrained use) and the fact that the study took place during the summer holidays. The similar number of participants who did not complete the study in the 2 groups (58% in the *Hygie* group and 62% in the control group) suggests that the reasons for not participating are not related to the nature of the teaching material. Similarly, it can be assumed that the influence of reminders during the week of access to teaching materials, compared with routine use, was similar for both groups. However, the final sample size was smaller than the number calculated as required. A lack of power may explain that 1 of the 2 end points did not reach statistical significance.

Knowledge and satisfaction questionnaire have not been previously validated because they have been made to match the content of the teaching materials. The use of customized instruments is strongly recommended for the evaluation of serious games by Moreno-Ger [37], who argues that generic questionnaires are usually not useful for assessing games that can be very different in their objectives, target audiences, and needs. However, GPs experienced in medical pedagogy reviewed and improved these questionnaires, which was then pilot tested.

The scores obtained by the participants in the pretest knowledge questionnaire were surprisingly low. The lack of knowledge of clinical practice recommendations by French GPs is known in the literature [38]. In addition, the knowledge questionnaire presented several difficulties: free-text responses and needing to know recent evidence-based recommendations. The improvement in scores between pretest and posttest, although significant in both groups, may appear small. In addition to the difficulty of the questionnaire, which may have limited the progression of participants, this slight increase can be explained by the forgetting of knowledge.

The duration between the end of access to the teaching materials and the final test questionnaire was chosen at 3 to 5 weeks to evaluate long-term memorization, the most relevant type of memorization for the GP, and to limit the number of people lost to follow-up over a too long a period. We based our decision of follow-up period on a study conducted in 2008 evaluating the long-term memorization by residents of recommendations on type 2 diabetes learned via an internet tutorial [39]. Subjects were randomized into 6 groups that varied the time between the tutorial and the knowledge assessment: without delay and with delay of 1 day, 3 days, 8 days, 21 days, and 55 days. At 21 days, the interns had forgotten more than half of the knowledge learned compared with those assessed without any delay, suggesting that this duration allows long-term learning to be assessed.

The time spent on learning material could not be collected automatically in control group. These data would have provided an additional element of comparison between the 2 groups. However, as the groups were randomized, the effect of individual preferences regarding time allocation can be assumed to be balanced between groups. The self-reported time of participants was more than 2 times longer in the *Hygie* group; it is possible that this result indicates that *Hygie* is more time-consuming than reading. However, in the "real life" conditions of this trial, where each participant chose the time spent on the support, it seems that this result is rather in favor of *Hygie*'s interest. That is, this educational support seems particularly engaging in this population of GPs, who are known to lack time and motivation for continuing education.

Conclusions

A very favorable reception was given by most GPs who used the *Hygie* game, particularly for its playful, interactive, and stimulating aspects, which supported the engaging learning experience.

In this study, many GPs spent much time on *Hygie*, commenting favorably on the clinical cases and the resulting learning experiences. A large proportion of participants expressed a desire to use it regularly for continuing education. In addition, *Hygie* serious game inspired many participants to subscribe to journals, which implies a synergy of this novel approach with the traditional article reading approach.

Our pragmatic study suggests that under usual conditions with e-learning teaching material, *Hygie* game can be an effective, pleasant, and engaging method for continuing education of GPs. It can be widely disseminated at low cost. Its modular content allows for future adaptation and improvement, and immersive qualities in a virtual reality where errors are not detrimental to patients render it an exciting next direction for adult learning among GPs and other physicians. In the future, we could evaluate the appropriation of this tool by GPs and their ability to improve it.

Acknowledgments

The authors would like to thank *Prescrire* journal, *Minerva* journal, French university departments, and resident supervisors for participating in this study; Prof Jean-Pierre Aubert, Prof Jean-François Bergmann, and Dr Josselin Le Bel for creating the

knowledge questionnaire; and Stéphanie Sidorkiewicz, Alexandre Malmartel, and Pascal Bessonneau for making corrections in the manuscript.

Conflicts of Interest

L-BJ is the owner of the *Hygie* (hygie-jeu.fr) website. No other COI declared.

Multimedia Appendix 1
CONSORT-EHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 393 KB - jmir_v21i11e12669_app1.pdf]

References

1. Kelly M, Murray TS. General practitioners' views on continuing medical education. Br J Gen Pract 1994 Oct;44(387):469-471 [FREE Full text] [Medline: 7748636]
2. Pitts J, Curtis A, While R, Holloway I. 'Practice professional development plans': general practitioners' perspectives on proposed changes in general practice education. Br J Gen Pract 1999 Dec;49(449):959-962 [FREE Full text] [Medline: 10824338]
3. ORS Pays de la Loire. General practitioners face the challenge of updating knowledge and practices and the issue of remuneration methods. Pays de la Loire: URML des Pays de la Loire; 2009.
4. Davis DA, Thomson MA, Oxman AD, Haynes RB. Evidence for the effectiveness of CME. A review of 50 randomized controlled trials. J Am Med Assoc 1992 Sep 2;268(9):1111-1117. [doi: 10.1001/jama.1992.03490090053014] [Medline: 1501333]
5. Grol R, Grimshaw J. From best evidence to best practice: effective implementation of change in patients' care. Lancet 2003 Oct 11;362(9391):1225-1230. [doi: 10.1016/S0140-6736(03)14546-1] [Medline: 14568747]
6. Mouaheb H, Fahli A, Moussetad M, Eljamali S. The serious game: what educational benefits? Procedia - Soc Behav Sci 2012;46:5502-5508. [doi: 10.1016/j.sbspro.2012.06.465]
7. Ricci KE, Salas E, Cannon-Bowers JA. Do computer-based games facilitate knowledge acquisition and retention? Mil Psychol 2009;8(4):295-307. [doi: 10.1207/s15327876mp0804_3]
8. Mayo MJ. Games for science and engineering education. Commun ACM 2007;50(7):30-35. [doi: 10.1145/1272516.1272536]
9. Larsen DP, Butler AC, Roediger HL. Repeated testing improves long-term retention relative to repeated study: a randomised controlled trial. Med Educ 2009 Dec;43(12):1174-1181. [doi: 10.1111/j.1365-2923.2009.03518.x] [Medline: 19930508]
10. Roediger HL, Butler AC. The critical role of retrieval practice in long-term retention. Trends Cogn Sci 2011 Jan;15(1):20-27. [doi: 10.1016/j.tics.2010.09.003] [Medline: 20951630]
11. Kerfoot BP, Baker H. An online spaced-education game for global continuing medical education: a randomized trial. Ann Surg 2012 Jul;256(1):33-38. [doi: 10.1097/SLA.0b013e31825b3912] [Medline: 22664558]
12. Asad M, Iqbal K, Sabir M. Effectiveness of problem based learning as a strategy to foster problem solving and critical reasoning skills among medical students. J Ayub Med Coll Abbottabad 2015;27(3):604-607. [Medline: 26721019]
13. Walljasper D. Games with goals. Nurse Educ 1982;7(1):15-18. [doi: 10.1097/00006223-198201000-00014] [Medline: 6915390]
14. Beylefeld AA, Struwig MC. A gaming approach to learning medical microbiology: students' experiences of flow. Med Teach 2007 Nov;29(9):933-940. [doi: 10.1080/01421590701601550] [Medline: 18158668]
15. Dankbaar M. Serious games and blended learning; effects on performance and motivation in medical education. Perspect Med Educ 2017 Mar;6(1):58-60 [FREE Full text] [doi: 10.1007/s40037-016-0320-2] [Medline: 27975195]
16. Allery LA. Educational games and structured experiences. Med Teach 2004 Sep;26(6):504-505. [doi: 10.1080/01421590412331285423] [Medline: 15763824]
17. Akl E, Sackett KM, Erdley WS, Mustafa RA, Fiander M, Gabriel C, et al. Educational games for health professionals. Cochrane Database Syst Rev 2013 Jan 31(1):CD006411. [doi: 10.1002/14651858.CD006411.pub3] [Medline: 23440807]
18. Ricciardi F, de Paolis LT. A comprehensive review of serious games in health professions. Int J Comput Games Technol 2014;2014:1-11. [doi: 10.1155/2014/787968]
19. Diehl LA, Souza RM, Alves JB, Gordan PA, Esteves RZ, Jorge ML, et al. InsuOnline, a serious game to teach insulin therapy to primary care physicians: design of the game and a randomized controlled trial for educational validation. JMIR Res Protoc 2013 Jan 21;2(1):e5 [FREE Full text] [doi: 10.2196/resprot.2431] [Medline: 23612462]
20. Diehl LA, Souza RM, Gordan PA, Esteves RZ, Coelho IC. InsuOnline, an electronic game for medical education on insulin therapy: a randomized controlled trial with primary care physicians. J Med Internet Res 2017 Mar 9;19(3):e72 [FREE Full text] [doi: 10.2196/jmir.6944] [Medline: 28279950]
21. Signoret J. SFMG Wealth Advisors. 2012. Evolution du contenu de la consultation de Médecine Générale en termes de maladies chroniques, aigues et de prises en charge non pathologiques entre 1993 et 2010 URL: http://www.sfmg.org/data/generateur/generateur_fiche/728/fichier_these_juliensignoretccb5a.pdf [accessed 2017-02-02]
22. La rédaction Prescrire. Hypertension artérielle essentielle chez l'adulte : premiers traitements. Rev Prescrire 2016 Jan:1-5.

23. La rédaction Prescrire. Traitement de l'hypertension arterielle essentielle : Un diuretique thiazidique en premier choix, le plus souvent. Rev Prescrire 2014 Apr:275.
24. Baeten R, Medica D. Preventive health checks: a health benefit? Minerva 2014;13(2):15-16 [FREE Full text]
25. La rédaction Prescrire. Coloscopies de dépistage: ne pas négliger les risques. Rev Prescrire 2015 Jun:453.
26. La rédaction Prescrire. Papillomavirus vaccines and Guillain-Barré syndromes: managing uncertainties. Rev Prescrire 2016;36(392):427-432 [FREE Full text]
27. La rédaction Prescrire. Patients ayant une hyperlipidémie. Rev Prescrire 2016;Guide Interactions médicamenteuses:159-167.
28. Niven D, Gaudet J, Laupland K. Brief analysis: are peripheral thermometers accurate enough to measure body temperature? Minerva 2016 May;Analyse brève:1-2 [FREE Full text]
29. Van den Bruel A, Haj-Hassan T, Thompson M. Diagnostic ambulatoire d?infections sévères chez les enfants. Minerva 2010 Dec;9(11):128-129.
30. La rédaction Prescrire. Rhume: Premiers Choix Prescrire, actualisation. Rev Prescrire 2015 Jul:1-4.
31. Jaunay L-B. Hygie game URL: http://hygie-jeu.fr/etude/ [accessed 2018-10-22] [WebCite Cache ID 73MXeHztv]
32. R: The R Project for Statistical Computing. URL: https://www.r-project.org/ [accessed 2018-08-20]
33. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. Br Med J 2010 Mar 23;340:c332 [FREE Full text] [doi: 10.1136/bmj.c332] [Medline: 20332509]
34. Bouton C, Leroy O, Huez J, Bellanger W, Ramond-Roquin A. Representativeness of General Practitioners university teachers. Public Health 2015;27(1):59-67. [doi: 10.3917/spub.151.0059]
35. La rédaction Prescrire. Les finances de l'Association Mieux Prescrire. Bilan de l'exercice 2015-2016. Rev Prescrire 2017:164-165.
36. Chevallier R. Banque de données en santé publique. 2017. Evaluation of an internal site for the management of type 2 diabetic patients, for use by general practitioners URL: http://bdsp-ehesp.inist.fr/vibad/index.php?action=getRecordDetail&idt=525817 [accessed 2019-07-26]
37. Moreno-Ger P, Torrente J, Hsieh YG, Lester WT. Usability testing for serious games: making informed design decisions with user data. Adv Hum-Comput Interact 2012;2012:1-13. [doi: 10.1155/2012/369637]
38. Morawski-Bachimont J, Cogneau J, Letourmy A. Why general practitioners do not respect recommendations for good clinical practices? The case of non insulin dependent diabetes mellitus. Sci Soc Santé 2006;24(2):75-103. [doi: 10.3406/sosan.2006.1681]
39. Bell DS, Harless CE, Higa JK, Bjork EL, Bjork RA, Bazargan M, et al. Knowledge retention after an online tutorial: a randomized educational experiment among resident physicians. J Gen Intern Med 2008 Aug;23(8):1164-1171 [FREE Full text] [doi: 10.1007/s11606-008-0604-2] [Medline: 18446414]

Abbreviations

DQ-5: Dynamic Questionnaire-5
e-learning: electronic learning
GP: general practitioner
PHP: Hypertext Preprocessor
SQ-20: Static Questionnaire-20

Edited by G Eysenbach; submitted 01.11.18; peer-reviewed by K Glover, J Kaewkungwal; comments to author 05.04.19; revised version received 17.06.19; accepted 19.07.19; published 20.11.19.

Please cite as:
Jaunay LB, Zerr P, Peguin L, Renouard L, Ivanoff AS, Picard H, Griffith J, Chassany O, Duracinsky M
Development and Evaluation of a New Serious Game for Continuing Medical Education of General Practitioners (Hygie): Double-Blinded Randomized Controlled Trial
J Med Internet Res 2019;21(11):e12669
URL: http://www.jmir.org/2019/11/e12669/
doi:10.2196/12669
PMID:31746775

Original Paper

Attracting Users to Online Health Communities: Analysis of LungCancer.net's Facebook Advertisement Campaign Data

Lindsey N Horrell[1*], BSN, MPH, PhD; Allison J Lazard[2,3*], PhD; Amrita Bhowmick[1,4*], MPH, MBA; Sara Hayes[4*], MPH; Susan Mees[4*], BA; Carmina G Valle[3,5*], MPH, PhD

[1]Department of Health Behavior, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States
[2]School of Media and Journalism, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States
[3]Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States
[4]Health Union, LLC, Philadelphia, PA, United States
[5]Department of Nutrition, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States
[*]all authors contributed equally

Corresponding Author:
Lindsey N Horrell, BSN, MPH, PhD
Department of Health Behavior
Gillings School of Global Public Health
University of North Carolina at Chapel Hill
170 Rosenau Hall, CB #7400
Chapel Hill, NC
United States
Phone: 1 502 644 7597
Email: horrell@email.unc.edu

Abstract

Background: With growing numbers of adults turning to the internet to get answers for health-related questions, online communities provide platforms with participatory networks to deliver health information and social support. However, to optimize the benefits of these online communities, these platforms must market effectively to attract new members and promote community growth.

Objective: The aim of this study was to assess the engagement results of Facebook advertisements designed to increase membership in the LungCancer.net online community.

Methods: In the fall of 2017, a series of 5 weeklong Facebook advertisement campaigns were launched targeting adults over the age of 18 years with an interest in lung cancer to increase opt ins to the LungCancer.net community (ie, the number of people who provided their email to join the site).

Results: The advertisements released during this campaign had a sum reach of 91,835 people, and 863 new members opted into the LungCancer.net community by providing their email address. Females aged 55 to 64 years were the largest population reached by the campaign (31,401/91,835; 34.29%), whereas females aged 65 and older were the largest population who opted into the LungCancer.net community (307/863; 35.57%). A total of US $1742 was invested in the Facebook campaigns, and 863 people opted into LungCancer.net, resulting in a cost of US $2.02 per new member.

Conclusions: This research demonstrates the feasibility of using Facebook advertising to promote and grow online health communities. More research is needed to compare the effectiveness of various advertising approaches. Public health professionals should consider Facebook campaigns to effectively connect intended audiences to health information and support.

(J Med Internet Res 2019;21(11):e14421) doi:10.2196/14421

KEYWORDS

internet; health communication; social media; health promotion; health education

Introduction

Online Community Growth

Currently, 72% of adults seek health information on the Web, and 16% search for peers with similar health concerns [1]. Online communities can effectively extend health education [2,3] and facilitate social support [3,4] and have been linked to improved self-management [2] and enhanced health outcomes [3]. The number of online communities has grown substantially over the past decade, with countless websites increasing traffic from patients and caregivers through user-engaged communities [5]. Patients are motivated to join these communities to access support, advice, and accountability in reaching health goals [5-7]. Online community growth is crucial to meeting these user needs, as it builds communities' pooled knowledge and increases access to quality informational and social support [5,8-11]. Larger online networks have the power of network effects—where more users increase the usefulness of the community [9]. For those seeking others with shared experiences, larger communities offer a greater number of individuals with the potential for cognitive empathy, particularly from people outside ones' close network where sharing may cause emotional burden [12]. For staff overseeing these sites, limited evidence is available to guide community growth, which is known to be a time- and resource-intensive task [8].

LungCancer.net

In this study, we reported the feasibility and cost-effectiveness of Facebook advertising to promote online community growth in the context of the LungCancer.net community. LungCancer.net provides patients and caregivers a platform to learn, educate, and connect with peers and health care professionals. The content published by LungCancer.net is written by patients, caregivers, and health professionals and supplemented by editorial content. In August 2017, LungCancer.net catered to 1575 users and sought to expand their community base through a series of social media advertisements. With 69% of US adults on Facebook and 74% of users on the site daily [13,14], Facebook seemed to be an ideal platform to promote community growth. The goal of this study was to assess the engagement results of Facebook advertisements designed to increase the number of opt ins to the LungCancer.net online community (ie, the number of users that provided their email to join the community).

Methods

Facebook Advertisement Campaign

From August to December 2017, 5 weeklong Facebook campaigns were launched with the objective of increasing opt ins to LungCancer.net. Each campaign consisted of 3 unique advertisements that contained an image, a text, and a call to action (Figure 1). The visuals included 6 static images and 1 image in the Graphics Interchange Format (signaled with the "†" symbol in Figure 1). The text included messages crafted by community managers and quotes from members. The target audience was adults (18 years or older) with an interest in lung cancer–related content and/or Facebook pages. No other demographic variables were used to define the audience within the Facebook Ads Manager system. The budget for each advertisement was US $25 per day. Facebook utilizes a bidding cost system, and actual expenditures for each test averaged within 4% of the desired budget, with the exception of 1 outlying test, which was 19% below the budget.

Advertisement Performance Measures

The performance of each advertisement was evaluated using metrics rooted in advertisement engagement frameworks [15-17]. According to McGuire's Model of Persuasion, eliciting action begins with advertisement exposure and moves across a continuum of cognitive and behavioral responses [17]. Exposure in this campaign is operationalized as *impressions* (number of times the advertisement appears in News Feeds) and *reach* (number of individuals exposed to the advertisement). Frameworks proposed by Neiger et al [15] and Platt et al [16] were used to define low-to-high behavioral responses. As the goal of this campaign was to increase opt ins to the LungCancer.net community, *low user engagement* was defined as interacting with the advertisement through clicks (ie, reacting to the post, clicking a post link, or liking the LungCancer.net Facebook page), *medium user engagement* was defined as sharing or commenting on the advertisement, and *high user engagement* was defined as *opting in* or signing up for the LungCancer.net community. After each campaign, metrics (Table 1) were pulled for each advertisement, and advertisements with the lowest opt in cost were run with new advertisements during the next weeklong campaign. Advertisements with the lowest opt in cost during each weeklong campaign are signaled with the "‡" symbol in Figure 1.

Figure 1. Facebook advertisement campaign images and text.

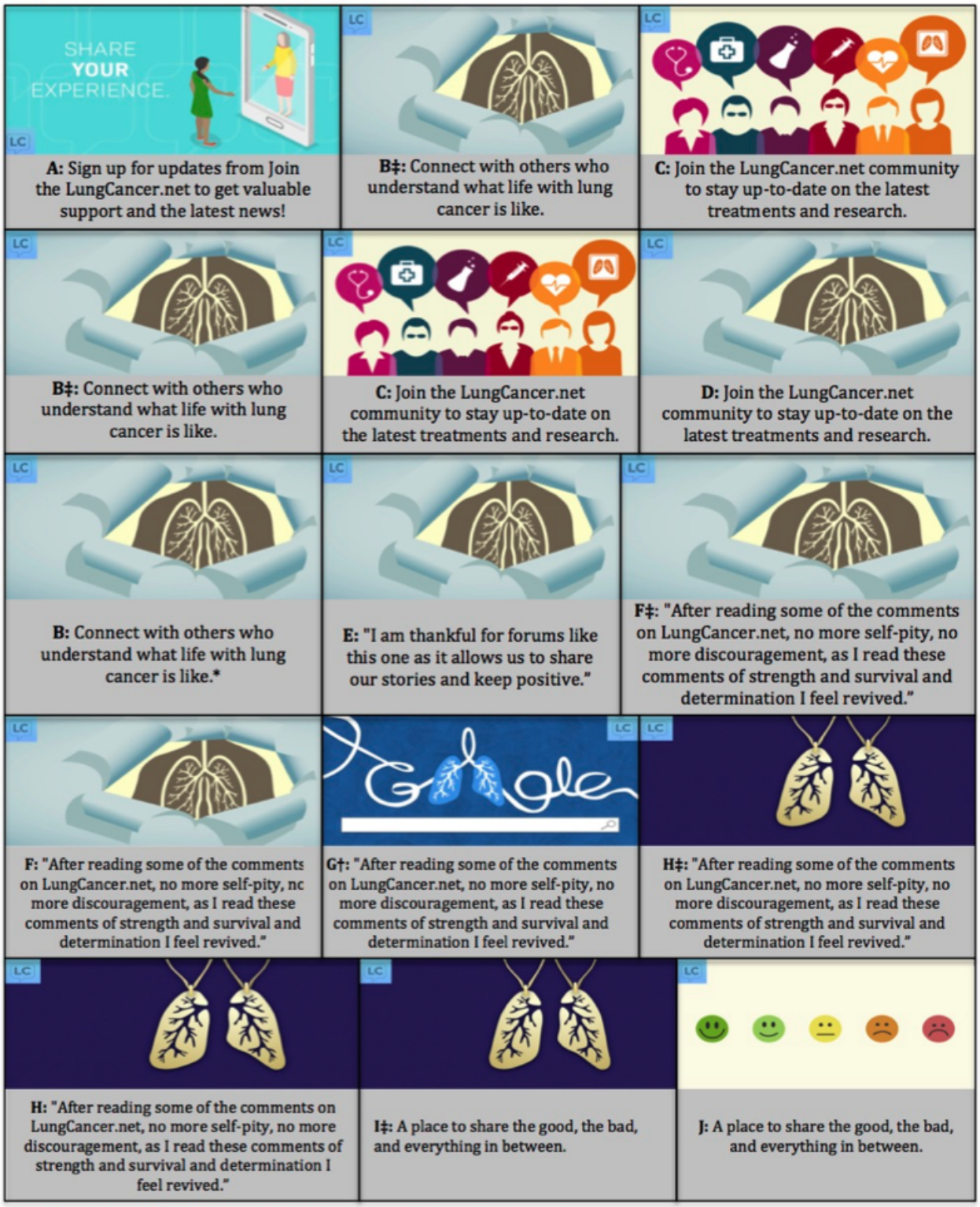


Table 1. Facebook advertisement performance measures.

Level of performance	Level of engagement	Performance measure	Definition
Exposure	—[a]	Impressions	Number of times the advertisement appeared in News Feeds
Exposure	—	Reach	Number of individuals exposed to the Facebook advertisement
Engagement	Low	Reactions	Number of times people responded to an advertisement by clicking “like,” “love,” “wow,” “haha,” “sad,” or “angry”
Engagement	Low	Link clicks	Number of people who clicked a link on the Facebook advertisement
Engagement	Low	Page likes	Number of people who liked the LungCancer.net Facebook page
Engagement	Medium	Shares	Number of times people shared the advertisement
Engagement	Medium	Comments	Number of times people commented on the Facebook advertisement
Engagement	High	Opt ins	Number of people who signed up to join the LungCancer.net community

[a]Not applicable.

Results

Audience Demographics

Over the course of the 5 campaigns, the sum reach was 91,835 people, and 863 members opted in to the LungCancer.net community (ie, demonstrated high engagement; Table 2). Females between 55 and 64 years represented the largest population reached by the campaign (31,401/91,835; 34.29%), whereas females aged 65 years and older represented the largest population that opted in to the LungCancer.net community (307/863; 35.57%). Given that US $1742 was invested across the 5 campaigns, approximately US $2.02 was spent per opt in, and just over 1 cent was spent per exposure to the campaign.

Advertisement Engagement Results

Table 3 displays engagement results. During the first campaign (August 24-30), advertisement B attracted the greatest level of engagement, including the greatest reach (10,556 people), number of impressions (12,569), reactions (221), link clicks (131), page likes (11), and opt ins (81 new community members) and the lowest opt in per cost rate (US $1.99 per opt in). This advertisement featured an image of lungs with the text “connect with others who understand what life with lung cancer is like.” Advertisements B and C were then used in the second campaign (August 31-September 6) alongside 1 new advertisement. In week 2, advertisement B again outperformed other advertisements and was subsequently implemented in week 3 (October 5-11).

During the third campaign, advertisement F, featuring the same image as advertisement B with new text “After reading some of the comments on LungCancer.net, no more self-pity, no more discouragement, as I read these comments of strength and survival and determination I feel revived” attracted the greatest number of reactions (194), comments (19), link clicks (176), and opt ins (82) at the lowest cost (US $1.10 per opt in). In the fourth campaign (November 9-15), advertisement H, with the same text as advertisement F but a simpler lung image, attracted the greatest engagement including 179 reactions, 14 page likes, and 60 opt ins at US $1.47 per opt in. In the fifth campaign (December 7-17), advertisement H was outperformed by an advertisement featuring the same image with the text, “A place to share the good, the bad, and everything in between” (advertisement I). Advertisement I attracted 114 link clicks, 22 page likes, and 50 opt ins at US $1.89 per opt in.

Table 2. Demographic information of those exposed to Facebook advertisements.

Age (years)	Cumulative campaign reach (N=91,835), n (%)			New members resulting from the campaign (N=863), n (%)		
	Female	Male	Unknown	Female	Male	Unknown
65+	24,005 (26.14)	5766 (6.28)	163 (0.18)	307 (35.57)	66 (7.65)	3 (0.35)
55-64	31,401 (34.29)	6572 (7.16)	181 (0.20)	257 (29.78)	63 (7.30)	3 (0.35)
45-54	13,289 (14.47)	2397 (2.61)	62 (0.07)	111 (12.86)	14 (1.62)	0 (0.00)
35-44	4427 (4.82)	975 (1.06)	20 (0.02)	23 (2.67)	1 (0.12)	0 (0.00)
25-34	138 (1.50)	404 (0.44)	12 (0.01)	7 (0.81)	2 (0.23)	0 (0.00)
18-24	591 (0.64)	115 (0.13)	15 (0.01)	6 (0.70)	0 (0.00)	0 (0.00)
Unknown	0 (0.00)	0 (0.00)	69 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)

Table 3. Facebook advertisement engagement results.

Ad	Exposure, n		Engagement, n						Cost/opt in rate (US $)
	Reach	Impressions	Low			Medium		High (opt ins)	
			Reactions	Link clicks	Page likes	Shares	Comments		
A	7206	8972	83	81	6	16	9	34	5.10
B[a]	10,556	12,569	221	131	11	44	6	81	1.99
C	8494	10,788	219	102	10	51	8	72	2.23
B[a]	10,546	12,965	170	164	9	34	15	78	1.85
C	9326	11,944	173	113	10	37	13	55	2.64
D	6484	9235	238	121	10	35	8	61	1.86
B	6078	8091	94	116	3	35	13	48	1.89
E	4018	6293	195	126	10	22	5	60	1.51
F[a]	4778	7262	194	176	9	26	19	82	1.10
F	4711	6313	151	134	10	33	11	60	1.73
G	3468	4530	141	75	7	23	5	35	2.53
H[a]	3952	5519	179	138	14	31	9	60	1.47
H	4671	5987	189	110	14	25	13	43	2.45
I[a]	4146	5472	171	114	22	23	8	50	1.89
J	3401	5067	184	88	4	17	15	44	2.12
Total	91,835	121,007	2602	1789	149	452	157	863	2.02

[a]Signals the highest performing ad (generated the most opt ins/cost) that was subsequently used in the next ad campaign.

Discussion

Principal Findings

Our findings demonstrate the feasibility of utilizing Facebook advertising as a cost-efficient tool to grow online health communities. Across the 5 campaigns, 863 new members opted in to the LungCancer.net community, yielding an opt in rate (opt ins/reach) of 0.94% (863/91,835) and a cost/opt in rate of US $2.02. Although the cost-effectiveness of Facebook advertisements varies widely in recruitment literature [18-24], our cost is but slightly higher than the average cost per click of US $1.32 for health care advertisements on Facebook [25]. Although Facebook advertisements were a cost-efficient community growth tool in this study, other research provides mixed results regarding the effectiveness of Facebook advertising [18,19,26,27]. Some agree that Facebook is an efficient way to draw diverse audiences to health promotion interventions [19,26,27]. Others have found Facebook to be a useful tool to increase advertisement reach, yet the actual rate of results per reach remains low [18,26]. This may indicate that Facebook advertisements are more efficient than traditional approaches (eg, physician referral, direct mail, and email) for online community growth outside research recruitment, where strict eligibility criteria often narrow the target audience [18]. Additional research is needed to test this hypothesis and optimize strategies to grow online health communities. Although these findings do not provide for specific design recommendations to increase engagement, we found some support for promising features of advertisements that match suggestions in previous literature: use of direct quotes/testimonials [28,29]; explicit reference to social support available in the community [6]; and simple lung images that are likely to be easily interpreted as relevant [30] to those seeking lung cancer communities.

Limitations and Future Research

Although this research provides foundational knowledge regarding the feasibility of Facebook advertisements to grow the LungCancer.net community, the findings are limited to the advertisement images and text used. Additional research is needed to systematically compare engagement with different images, texts, channels, and times of year to identify strategies associated with optimal community growth. Research is also needed to identify the impact that community growth through Facebook advertisements has on community engagement. Users who respond to a Facebook advertisement already demonstrate online engagement and may be more likely to contribute to an online health community than members recruited through other traditional strategies. Finally, given suggestions that Facebook advertising can effectively engage hardly reached populations in health education and intervention [15,18-20,27,31,32], additional research is needed to identify the sociodemographic characteristics of those engaged. Data presented here demonstrate a campaign that engaged primarily ageing female populations, representative of the current LungCancer.net site visitors (61% female and 55 years and above).

Conclusions

This study provides a foundation for research to optimize the reach of online health communities. Facebook was a feasible, cost-effective recruitment channel for this online community, and evaluation of other advertisement designs may provide further evidence for promising engagement strategies. Online communities are vital to health promotion efforts as multiple populations seek low-cost, easily accessible health resources. Focusing on expanding the reach of such communities could have major implications for the health of future populations.

Acknowledgments

Research reported in this publication was supported by the National Cancer Institute of the National Institutes of Health under Award Number T32CA128582.

Conflicts of Interest

AB and SH are employees of Health Union, LLC.

References

1. Fox S. Pew Research Center. 2014. The Social Life of Health Information URL: http://www.pewresearch.org/fact-tank/2014/01/15/the-social-life-of-health-information/ [accessed 2019-09-17]
2. Willis E. The making of expert patients: the role of online health communities in arthritis self-management. J Health Psychol 2014 Dec;19(12):1613-1625. [doi: 10.1177/1359105313496446] [Medline: 23988679]
3. Wright KB. Communication in health-related online social support groups/communities: a review of research on predictors of participation, applications of social support theory, and health outcomes. Rev Commun Res 2016;4:65-87 [FREE Full text] [doi: 10.12840/issn.2255-4165.2016.04.01.010]
4. Moon TJ, Chih MY, Shah DV, Yoo W, Gustafson DH. Breast cancer survivors' contribution to psychosocial adjustment of newly diagnosed breast cancer patients in a computer-mediated social support group. Journalism Mass Commun Q 2017 Jan 19;94(2):486-514 [FREE Full text] [doi: 10.1177/1077699016687724]
5. Malanga E, Malanga V, Latham S. Websites, online communities and digital channels as a medium to effectively educate, engage and empower patients. Chronic Obstr Pulm Dis 2018 Oct 17;5(4):334-337 [FREE Full text] [doi: 10.15326/jcopdf.5.4.2018.0158] [Medline: 30723789]
6. Solberg LB. The benefits of online health communities. Virtual Mentor 2014 Apr 1;16(4):270-274 [FREE Full text] [doi: 10.1001/virtualmentor.2014.16.04.stas1-1404] [Medline: 24735575]
7. Newman MW, Lauterbach D, Munson SA, Resnick P, Morris ME. It's Not That I Don't Have Problems, I'm Just Not Putting Them on Facebook: Challenges and Opportunities in Using Online Social Networks for Health. In: Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work. 2011 Presented at: CSCW'11; March 19-23, 2011; Proceedings of the Acm 2011 Conference on Computer Supported Cooperative Work p. 341-350 URL: https://tinyurl.com/y6trecr2 [doi: 10.1145/1958824.1958876]
8. Young C. Community management that works: how to build and sustain a thriving online health community. J Med Internet Res 2013 Jun 11;15(6):e119 [FREE Full text] [doi: 10.2196/jmir.2501] [Medline: 23759312]
9. Lin KY, Lu HP. Why people use social networking sites: an empirical study integrating network externalities and motivation theory. Comput Hum Behav 2011 May;27(3):1152-1161. [doi: 10.1016/j.chb.2010.12.009]
10. Akar E, Mardikyan S. User roles and contribution patterns in online communities: a managerial perspective. Sage Open 2018 Aug 22;8(3):215824401879477. [doi: 10.1177/2158244018794773]
11. Butler BS. Membership size, communication activity, and sustainability: a resource-based model of online social structures. Inf Syst Res 2001 Dec;12(4):346-362. [doi: 10.1287/isre.12.4.346.9703]
12. Small ML. Weak ties and the core discussion network: why people regularly discuss important matters with unimportant alters. Social Networks 2013 Jul;35(3):470-483. [doi: 10.1016/j.socnet.2013.05.004]
13. Pew Research Center. 2019. Social Media Fact Sheet URL: https://www.pewinternet.org/fact-sheet/social-media/ [accessed 2019-09-17]
14. Smith A, Anderson M. Pew Research Center. 2018. Social Media Use in 2018 URL: https://www.pewinternet.org/2018/03/01/social-media-use-in-2018/ [accessed 2019-09-17]
15. Neiger BL, Thackeray R, van Wagenen SA, Hanson CL, West JH, Barnes MD, et al. Use of social media in health promotion: purposes, key performance indicators, and evaluation metrics. Health Promot Pract 2012 Mar;13(2):159-164. [doi: 10.1177/1524839911433467] [Medline: 22382491]
16. Platt T, Platt J, Thiel DB, Kardia SL. Facebook advertising across an engagement spectrum: a case example for public health communication. JMIR Public Health Surveill 2016 May 30;2(1):e27 [FREE Full text] [doi: 10.2196/publichealth.5623] [Medline: 27244774]
17. McGuire WJ. Theoretical foundations of campaigns. In: Rice RE, Atkin CK, editors. Public Communication Campaigns. Second Edition. Newbury Park, CA: Sage Publications; 1989:43-67.

18. Frandsen M, Thow M, Ferguson SG. The effectiveness of social media (Facebook) compared with more traditional advertising methods for recruiting eligible participants to health research studies: a randomized, controlled clinical trial. JMIR Res Protoc 2016 Aug 10;5(3):e161 [FREE Full text] [doi: 10.2196/resprot.5747] [Medline: 27511829]
19. Lane TS, Armin J, Gordon JS. Online recruitment methods for web-based and mobile health studies: a review of the literature. J Med Internet Res 2015 Jul 22;17(7):e183 [FREE Full text] [doi: 10.2196/jmir.4359] [Medline: 26202991]
20. Chung SY, Hacker ED, Rawl S, Ellis R, Bakas T, Jones J, et al. Using Facebook in recruiting kidney transplant recipients for a REDCap study. West J Nurs Res 2019 Mar 5:- (epub ahead of print). [doi: 10.1177/0193945919832600] [Medline: 30836840]
21. Guthrie KA, Caan B, Diem S, Ensrud KE, Greaves SR, Larson JC, et al. Facebook advertising for recruitment of midlife women with bothersome vaginal symptoms: a pilot study. Clin Trials 2019 Oct;16(5):476-480. [doi: 10.1177/1740774519846862] [Medline: 31055949]
22. Fenner Y, Garland SM, Moore EE, Jayasinghe Y, Fletcher A, Tabrizi SN, et al. Web-based recruiting for health research using a social networking site: an exploratory study. J Med Internet Res 2012 Feb 1;14(1):e20 [FREE Full text] [doi: 10.2196/jmir.1978] [Medline: 22297093]
23. Akard TF, Wray S, Gilmer MJ. Facebook advertisements recruit parents of children with cancer for an online survey of web-based research preferences. Cancer Nurs 2015;38(2):155-161 [FREE Full text] [doi: 10.1097/NCC.0000000000000146] [Medline: 24945264]
24. Thornton LK, Harris K, Baker AL, Johnson M, Kay-Lambkin FJ. Recruiting for addiction research via Facebook. Drug Alcohol Rev 2016 Jul;35(4):494-502. [doi: 10.1111/dar.12305] [Medline: 26179224]
25. Irvine M. WordStream: Online Advertising Made Easy. 2019. Facebook Ad Benchmarks for Your Industry URL: https://www.wordstream.com/blog/ws/2017/02/28/facebook-advertising-benchmarks [accessed 2019-09-17]
26. Topolovec-Vranic J, Natarajan K. The use of social media in recruitment for medical research studies: a scoping review. J Med Internet Res 2016 Nov 7;18(11):e286 [FREE Full text] [doi: 10.2196/jmir.5698] [Medline: 27821383]
27. Pedersen ER, Kurz J. Using Facebook for health-related research study recruitment and program delivery. Curr Opin Psychol 2016 May;9:38-43 [FREE Full text] [doi: 10.1016/j.copsyc.2015.09.011] [Medline: 26726313]
28. Davis KC, Nonnemaker JM, Farrelly MC, Niederdeppe J. Exploring differences in smokers' perceptions of the effectiveness of cessation media messages. Tob Control 2011 Jan;20(1):26-33. [doi: 10.1136/tc.2009.035568] [Medline: 20852323]
29. de Wit JBF, Das E, Vet R. What works best: objective statistics or a personal testimonial? An assessment of the persuasive effects of different types of message evidence on risk perception. Health Psychol 2008 Jan;27(1):110-115. [doi: 10.1037/0278-6133.27.1.110] [Medline: 18230021]
30. Petty RE, Cacioppo JT, Schumann D. Central and peripheral routes to advertising effectiveness: the moderating role of involvement. J Consum Res 1983 Sep;10(2):135. [doi: 10.1086/208954]
31. Saha K, Weber I, Birnbaum ML, de Choudhury M. Characterizing awareness of schizophrenia among Facebook users by leveraging Facebook advertisement estimates. J Med Internet Res 2017 May 8;19(5):e156 [FREE Full text] [doi: 10.2196/jmir.6815] [Medline: 28483739]
32. Dunn PH, Woo BK. Facebook recruitment of Chinese-speaking participants for hypertension education. J Am Soc Hypertens 2018 Sep;12(9):690-692. [doi: 10.1016/j.jash.2018.06.017] [Medline: 30054194]

Edited by G Eysenbach; submitted 20.04.19; peer-reviewed by T Platt, SK Ernala, E Da Silva, P Matthews; comments to author 21.06.19; revised version received 23.08.19; accepted 30.08.19; published 04.11.19.

Please cite as:
Horrell LN, Lazard AJ, Bhowmick A, Hayes S, Mees S, Valle CG
Attracting Users to Online Health Communities: Analysis of LungCancer.net's Facebook Advertisement Campaign Data
J Med Internet Res 2019;21(11):e14421
URL: https://www.jmir.org/2019/11/e14421
doi:10.2196/14421
PMID:31682589

Original Paper

Effects of Three Antecedents of Patient Compliance for Users of Peer-to-Peer Online Health Communities: Cross-Sectional Study

Anne-Françoise Audrain-Pontevia[1], PhD; Loick Menvielle[2], PhD; Myriam Ertz[3], PhD

[1]École des Sciences de la Gestion, Université du Québec à Montréal, Montréal, QC, Canada

[2]École des Hautes Études Commerciales du Nord, Nice, France

[3]Université du Québec à Chicoutimi, Chicoutimi, QC, Canada

Corresponding Author:
Anne-Françoise Audrain-Pontevia, PhD
École des Sciences de la Gestion
Université du Québec à Montréal
315 rue Sainte Catherine Est
Montréal, QC, H2X 3X2
Canada
Phone: 1 514 489 3000 ext 3572
Email: audrain_pontevia.anne_francoise@uqam.ca

Abstract

Background: Over the past 50 years, patient noncompliance has appeared as a major public health concern and focus of a great deal of research because it endangers patient recovery and imposes a considerable financial burden on health care systems. Meanwhile, online health communities (OHCs) are becoming more common and are commonly used by individuals with health problems, and they may have a role in facilitating compliance. Despite this growing popularity, little is known about patient compliance predictors for OHCs' users.

Objective: This study aimed to investigate the extent to which participating in OHCs may trigger higher levels of compliance. It identified 3 interrelated predictors that may affect patient compliance: patient empowerment gained through peer-to-peer OHCs, satisfaction with the physician, and commitment to the physician.

Methods: A Web-based survey tested the conceptual model and assessed the effects of patient empowerment gained through OHCs on patient satisfaction and commitment to the physician, as well as the effects of these 3 predictors on patient compliance with the proposed treatment. Members of peer-to-peer OHCs were asked to answer an online questionnaire. A convenience sample of 420 patients experiencing chronic illness and using peer-to-peer OHCs was surveyed in August 2018 in Québec, Canada. A path analysis using structural equation modeling tested the proposed relationships between the predictors and their respective paths on patient compliance. The mediation effects of these predictor variables on patient compliance were estimated with the PROCESS macro in SPSS.

Results: The findings indicated that patient empowerment gained through OHCs was positively related to patient commitment to the physician (beta=.69; P<.001) and patient compliance with the proposed treatment (beta=.35; P<.001). Patient commitment also positively influenced patient compliance (beta=.74; P<.001). Patient empowerment did not exert a significant influence on patient satisfaction with the physician (beta=.02; P=.76), and satisfaction did not affect compliance (beta=−.07; P=.05); however, patient satisfaction was positively related to patient commitment to the physician (beta=.14; P<.01). The impact of empowerment on compliance was partially mediated by commitment to the physician (beta=.32; 95% CI 0.22-0.44) but not by satisfaction.

Conclusions: This study highlights the importance of peer-to-peer OHCs for two main reasons. The primary reason is that patient empowerment gained through peer-to-peer OHCs both directly and indirectly enhances patient compliance with the proposed treatment. The underlying mechanisms of these effects were shown. Second, commitment to the physician was found to play a more critical role than satisfaction with the physician in determining patient-physician relationship quality. Overall, our findings support the assumption that health care stakeholders should encourage the use of peer-to-peer OHCs to favor patient empowerment and patient commitment to the physician to increase patient compliance with the proposed treatment.

(J Med Internet Res 2019;21(11):e14006) doi:10.2196/14006

KEYWORDS

online social networking; patient empowerment; patient compliance; patient satisfaction; structural equation modeling

Introduction

Context

Defined as the extent to which a patient's behavior coincides with the medical or health advice given by a health care specialist, patient compliance is of particularly critical importance for people with chronic health problems [1-3]. In developed countries, the World Health Organization [4] estimates that only 50% of chronically ill patients follow their prescribed treatment. Numerous empirical studies have aimed at describing and understanding patient compliance and corollary noncompliance over the past decades [5]. The literature emphasizes that patients' lack of compliance with prescribed therapeutic regimens jeopardizes their health, adversely affects treatment outcomes, and leads to wasted health care resources [6]. Prior research has shown that nonadherence may cause 125,000 avoidable deaths each year and cost US $100 billion annually in preventable health care expenditures [7]. By seeking to unveil the factors associated with lack of compliance, scholars identified demographic and cultural differences as well as psychological or social factors, among others [8]. However, the literature highlights a dearth of consistencies and consensus regarding the determinants of patient compliance.

As increasing patient compliance is estimated to be more critical to improving the health of a population than any advancement in medical treatment [4,9], it is of strategic interest to understand the determinants of patient compliance, particularly in the era of the *medical internet*. Today's digitization of health care holds promising perspectives for improving patient commitment and compliance [10]. Online health communities (OHCs) emphasize user-generated content and make it possible for users to exchange medical information anonymously, with no temporal or geographical constraints. OHCs are small virtual discussion groups in which people with a common concern about a health topic share information, experiences, and feelings; provide advice to fellow members; and provide social and emotional support [11,12]. Although these communities can present disadvantages, such as the spread of misinformation and unreliable support or advice, OHCs play a role in heightening patients' sense of empowerment as they feel better informed and guided by relevant others [13]. Prior research has shown that communication between the physicians and patients in these communities enhances patient compliance [14]. However, despite the considerable development of peer-to-peer OHCs, there is a lack of evidence regarding their effects on patient compliance.

This study examined the relationships between 3 predictors that are theoretically and nomologically related to patient compliance, for patients who are active on peer-to-peer OHCs. These predictors were patient empowerment, patient satisfaction with their physician, and patient commitment to the relationship with the physician. We analyzed these constructs because they have been identified as critical predictors of patient compliance [15,16]. Previous research has shown that patient empowerment predicts patient commitment to the relationship with the physician [17]. In this study, we focused on patient compliance with recommended treatment, that is, the extent to which the patient adheres to prescriptions and treatment recommendations targeted to his or her disease. We intended to answer the following research questions: what are the antecedents of patient compliance with the treatment?, how does empowerment impact patient compliance with the treatment?, and what is the mechanism underlying the impact of empowerment on patient compliance with the treatment?. This study may equip physicians, other health care professionals, scholars, managers, and decision makers alike, with improved insights into a promising and original avenue for encouraging chronically ill patients to comply with their treatment.

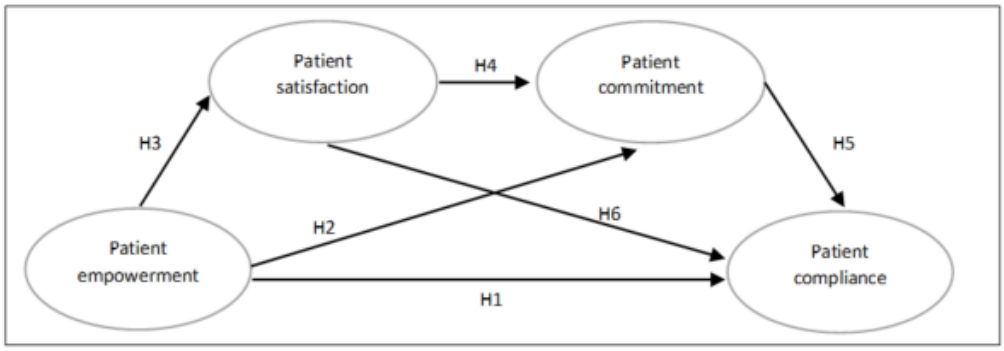
Theoretical Background and Model

Given the growing importance of a patient-centric perspective in contemporary health care, patient empowerment has been the subject of considerable attention from researchers over the past decades [18,19]. This is a consequence of the focus on a perspective that considers patients as consumers who also increasingly embrace the self-management of their disease [20,21]. Despite the amount of research dedicated to this construct in the health care research domain, there is a surprising lack of a consensual definition [22,23]. Most definitions agree that patient empowerment refers to the enhanced ability of patients to understand and influence their health [24]. Patient empowerment is frequently conceived of as a multidimensional and cognitive concept that includes different competencies and skills [25]. It refers to patients' control over their illness and or treatment as well as to their ability to understand and participate in the consultation and contribute to the decision process based on the support brought by the physician [11,17,24].

There appears to be a broad consensus on the virtuous effects of empowerment [26]. Empowered patients tend to be more in control of their disease because it helps them to reduce uncertainty and to develop better strategies to cope with the disease [27]. Patient empowerment may also enhance patient commitment toward the physician [17] and patient satisfaction with the physician [28]. Besides, patient empowerment is identified as a prerequisite for efficient patient-physician relationships [29], and it increases patient compliance with physician-proposed treatment [17,30]. It appears that patient empowerment leads to higher levels of self-efficacy and self-management, which in turn result in better outcomes such as patient quality of life while reducing costs for health care systems. For all these reasons, patient empowerment has become a priority for health care systems dealing with chronic diseases [21,31].

Building on previous research in the health care and consumer research fields [17,30,32], we proposed a conceptual model relating patient empowerment gained through peer-to-peer OHCs and patient compliance with physician-recommended treatment (Figure 1).

Figure 1. Theoretical model of the predictors of patient compliance showing hypothesized (H) relationships.



Specifically, we hypothesized the following:

H1: Patient empowerment gained through peer-to-peer OHC would positively relate to patient compliance with the proposed treatment.

H2: Patient empowerment gained through peer-to-peer OHC would positively relate to patient commitment to the physician.

H3: Patient empowerment gained through peer-to-peer OHC would positively relate to patient satisfaction toward the physician.

In addition to the abovementioned direct relationships, we expected that the relationship between patient empowerment and patient compliance would be mediated by patient satisfaction with the physician and patient commitment toward the physician (Figure 1). These hypotheses relied on the relationship literature positing that both consumer satisfaction and commitment are prerequisites to a consumer-firm relationship. Hence, we hypothesized the following:

H4: Patient satisfaction with the physician would positively relate to patient commitment toward the physician.

H5: Patient commitment toward the physician would positively relate to patient compliance with the proposed treatment.

H6: Patient satisfaction with the physician would positively relate to patient compliance with the proposed treatment.

Purpose and Contributions

Overall, 6 research hypotheses were tested on a total sample of 420 chronically ill OHC members by combining both structural equation modeling (SEM) and the bootstrapping-based PROCESS macro in SPSS (n=10,000 sample replications) to analyze the interrelationships within the proposed conceptual framework grounded in the fields of psychology, health care, and the social sciences. SEM has the particularity of testing and estimating simultaneously complex causal relationships between latent variables that are operationalized by manifest variables. Besides, SEM estimates random errors in the observed variables, thus reinforcing estimation accuracy. To assess the existence of mediating effects, we used the PROCESS macro in SPSS. This research contributes to the existing literature in several ways. First, it proposed a conceptual framework relating patient empowerment gained through peer-to-peer OHCs to patient compliance. Second, it tested both direct and indirect effects within patient empowerment gained through peer-to-peer OHCs in a framework of nomologically related variables. As such, the study highlighted the key contributors to patient compliance for chronically ill patients involved in peer-to-peer OHCs. Third, through the study's focus on chronically ill patients, this research provides valuable insights to better support a population that is most vulnerable to death and disability worldwide [21,33].

Methods

Participant Recruitment and Data Collection

Data for this cross-sectional study were collected over a 3-week period in August 2018 from OHCs in Canada. A self-reported questionnaire was administered on the Web-based survey platform Qualtrics. The link to the questionnaire was posted on Canadian OHCs dedicated to chronic diseases for a representative sample from the Qualtrics panel using preset quotas. We focused on French-speaking peer-to-peer OHCs on which members report suffering from at least one chronic sickness such as, for example, cancer, diabetes, or obesity. Participants had to have visited an OHC at least once during the 3 months before the survey. The questionnaire was pretested (10 respondents) and pilot-tested (32 respondents) because of the translation and adaptation of items to French and to confirm validity and reliability. Informed consent was obtained from each participant before starting the survey. Participation was voluntary and anonymous. An ethics committee has validated and approved the research protocol.

Sample Size

The study size was estimated based on a rule of thumb of roughly 10 respondents per item in the study [34]. With a total of 14 items, a minimum sample size of 140 observations was required. An initial pool of 1760 respondents entered the survey; 420 of these respondents matched the inclusion criteria. After removing incomplete or invalid questionnaires, the final operative sample was composed of 315 observations.

Results

Characteristics of the Study Population

Demographic characteristics of the sample are shown in Table 1. Of the 315 respondents, 144 (45.7%) were female and 134

(42.5%) were younger than 35 years. In addition, 208 out of 305 (68.2%) of the respondents had a university degree. Within this chronically ill patient sample, 132 out of 306 (43.1%) respondents suffered from type 1 or type 2 diabetes; thus, diabetes was the most represented pathology. Furthermore, 66 out of 306 (21.6%) respondents declared that they suffered from obesity. Regarding the duration of the illness, most of the respondents, 236 out of 315 (75%), declared that they had suffered from a chronic illness for at least 1 year.

Table 1. Demographic characteristics of study participants (N=315).

Demographic characteristics	Value, n (%)
Gender	
Female	144 (45.7)
Male	171 (54.3)
Age (years)	
Less than 18	6 (1.9)
18-24	45 (14.3)
25-34	70 (22.2)
35-49	110 (34.9)
50-65	73 (23.2)
More than 65	11 (3.5)
Education	
No education	30 (9.5)
Secondary	67 (21.3)
Undergraduate	84 (26.7)
Graduate	49 (15.6)
Postgraduate	75 (23.8)
Missing data	10 (3.1)
Chronic disease	
Diabetes type 1	62 (19.7)
Diabetes type 2	70 (22.2)
Obesity	66 (21.6)
HIV	3 (1)
Cancer	15 (4.7)
Other diseases	90 (28.5)
Missing data	9 (2.9)
Duration of the illness (years)	
<1	78 (24.8)
1 to <2	138 (43.8)
2 to <3	64 (20.3)
>3	34 (10.8)
Missing data	1 (0.3)

Measurement of Variables

As our research investigated OHCs from consumer behavior and marketing perspectives, we have grounded our measurements in both the behavioral and psychological theories as well as measurement tools. For the measure of empowerment, we used an adapted version of Ouschan et al's validated scale [17], consisting of 15 items, reduced to 4 items (mean 3.28, SD 1.58) after the factor analysis revealed several inconsistencies with this scale (see details in the next subsection). We measured patient satisfaction by adapting Oliver's scale [35], consisting of 3 items (mean 3.83, SD 1.79). Patient commitment was measured by 4 items from Morgan and Hunt's commitment scale (mean 3.67, SD 1.91) [36]. Finally, patient compliance, our dependent variable, was captured with 3 items from Prigge et al's compliance scale (mean 3.10, SD 1.62) [16]. All items

were scored on 7-point Likert scales, ranging from 1 (*totally disagree*) to 7 (*totally agree*). The items retained at the end of the factorial analysis purification procedure are shown in Table 2. We used gender, age, nationality, province or country of residence, education, and occupation as control variables. To reduce measurement context effects and common method bias (CMB), the measurement items were randomized within the research questionnaire.

Measurement of the Research Model

A set of preliminary analyses such as outliers, nonnormality checks, and descriptive data analysis was carried out. To check for CMB inherent to cross-sectional survey-based studies, Harman single factor test was run by performing factor analysis. The results revealed that the first factor does not explain more than 50% of the overall variance, which showed that CMB was not a concern. To test the measurement model, an exploratory factor analysis with the principal component analysis extraction method and the varimax rotation technique was carried out. It showed that the measurement structure explained 83.8% of the variance and that each item loaded significantly on its intended factor.

To assess the fit of the measurement model, a confirmatory factor analysis (AMOS in SPSS) was performed with the maximum likelihood estimation procedure, a robust method in latent variable modeling [37]. Different goodness-of-fit indices were used to estimate the quality of the model, namely, comparative fit index (CFI), normed fit index (NFI), nonnormed fit index (NNFI) greater than or equal to 0.950, standardized root mean residual (SRMR) less than or equal to 0.080, and root mean square error of approximation (RMSEA) less than or equal to 0.050 [38]. However, the fit of the model was consistently poor because of the empowerment items. This is not surprising as there is a lack of consensus on both the conceptualization and the measurement of empowerment [39,40]. According to the Construct definition, Object classification, Attribute classification, Rater identification, Scale formation, and Enumeration and reporting methodology, the construct appears thus misspecified, and the measurement tools supposed to measure this concept might actually measure something else [41]. Therefore, after several iterations, a total of 11 items were deleted from the empowerment scale. The resulting measurement model displayed good overall fit (χ^2_{36}=43.5 CFI=0.997, NFI=0.983, NNFI=0.995, SRMR=0.026, and RMSEA=0.028). Besides, all item loadings were significant and above the 0.70 threshold, and, as shown in Table 2, the average variance extracted (AVE) of each construct (from 0.644 to 0.799) was above the 0.50 threshold. Conjointly, these results demonstrate convergent validity [38]. As further shown in Table 3, the coefficient of reliability was higher than the AVE for each construct, thus reinforcing convergent validity [42].

Table 2. Constructs, items, means, standard loading, and standard deviation.

Construct and item	Standard loading	Mean (SD)
Empowerment (physician support)		
When addressing my condition, my doctor focuses on health promotion	0.69	3.21 (1.57)
My doctor provides clear instructions on what to do in different situations	0.78	3.27 (1.51)
When appropriate, my doctor provides me with a written plan on how to control my chronic illness condition	0.82	3.33 (1.61)
My doctor keeps me up to date with the most recent information on chronic illness conditions	0.83	3.32 (1.60)
Commitment		
The relationship with my physician is important for me	0.85	3.00 (1.59)
The relationship with my physician is something that I want to maintain	0.87	3.08 (1.50)
The relationship with my physician is particularly important for me	0.77	3.25 (1.59)
Satisfaction		
I am satisfied with my physician	0.92	3.67 (1.91)
I think I made the good choice to choose my physician	0.93	3.90 (1.78)
If I had to choose a physician, I would choose another one (reversed-polarity item)	0.92	3.92 (1.69)
Compliance		
I take the medication prescribed by my doctor at the right time	0.62	3.40 (1.67)
I take the right dosage of the medication prescribed by my doctor	0.76	2.84 (1.64)
I follow the prescribed treatment regularly and continuously	0.66	3.06 (1.56)

Table 3. Psychometric properties.

Construct	Cronbach alpha	Average variance extracted	Average shared variance	Composite reliability	Empowerment	Satisfaction	Commitment	Compliance
Empowerment	.883	0.719	0.408	0.885	0.848	0.001	0.719	0.840
Satisfaction	.924	0.799	0.007	0.923	0.001	*0.894*[a]	0.139	0.030
Commitment	.895	0.741	0.425	0.895	0.719	0.139	*0.861*	0.859
Compliance	.779	0.644	0.481	0.783	0.840	0.030	0.859	*0.802*

[a]The italicized values in the diagonal refer to the average variance extracted (AVE) for each construct.

Pearson product-moment correlation analyses revealed both univariate and bivariate links between the variables. As suggested by Fornell and Larcker, each latent variable accounted for more variance, as shown in its AVE, than its shares with the other constructs in the model, as shown in the interconstruct correlations, except for compliance [42]. Although both correlations were below the maximum tolerable threshold of 0.90 for a correlation, we checked that compliance was unrelated to both empowerment (r=0.84) and commitment (r=0.85) at the .01 level, for which high correlations were found [43]. We constructed a confidence interval for both correlations (ie, correlation ± 1.96 x standard error [44]. If |1| is included in this interval, this indicates a lack of discriminant validity [44]. The resulting confidence intervals for the empowerment, compliance correlation (0.84 ± 1.96 x standard error 0.06; CI 0.72-0.95) as well as for the compliance, commitment correlation (0.85 ± 1.96 x standard error 0.06; CI 0.73-0.98) both excluded |1|, confirming discriminant validity. Both the coefficients of reliability (from 0.885 to 0.923) and Cronbach alpha (from .883 to .924) were high, indicating good construct reliability [42].

Structural Model

We evaluated the fit of our structural model. The results suggest that the collected data adequately fit our research model as an appropriate model fit has been reached (χ^2_{35}=43.5; CFI=0.997, NFI=0.983, NNFI=0.995, SRMR=0.026, and RMSEA=0.028). The path results are visually summarized in Figure 2.

Overall, the results show that all the anticipated relationships of our model were supported except for the impact of empowerment on satisfaction and satisfaction on compliance. Empowerment was positively related to compliance, lending support to hypothesis 1 but suggesting a partial mediation effect of patient satisfaction and patient commitment on the relationship between patient empowerment and patient compliance. Empowerment positively influenced patient commitment, supporting hypothesis 2. However, patient satisfaction did not appear to play a significant role in explaining the effect of empowerment gained through OHCs and patient compliance with the prescribed treatment, as the relationship between patient empowerment and satisfaction was nonsignificant, as was the link between patient satisfaction and patient compliance. These results invalidate hypotheses 3 and 6, respectively. These findings suggest that the primary explanatory variable in the relationship was patient commitment as patient empowerment significantly increased the level of patient commitment. In turn, commitment improves compliance with the prescribed treatment. Collectively, these findings support hypotheses 2 and 5. The results provide preliminary evidence that patient commitment partially explains the relationship between empowerment gained through OHCs and enhanced patient compliance, whereas satisfaction does not (see Table 4).

Figure 2. Structural model with results. ns: not significant.

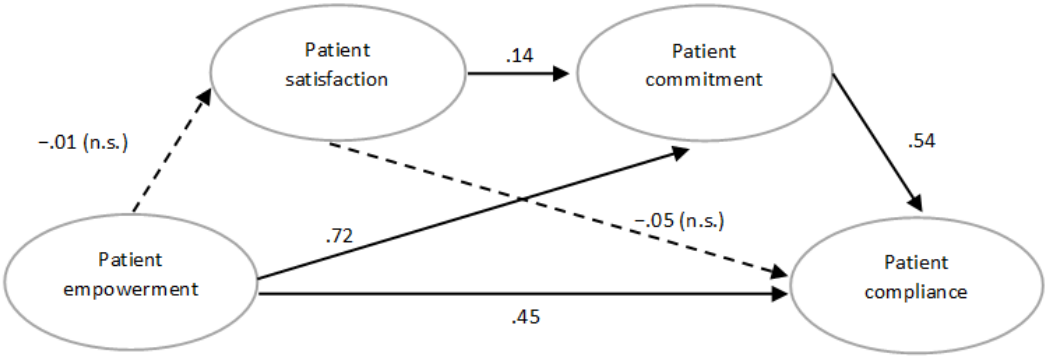


Table 4. Standardized coefficients.

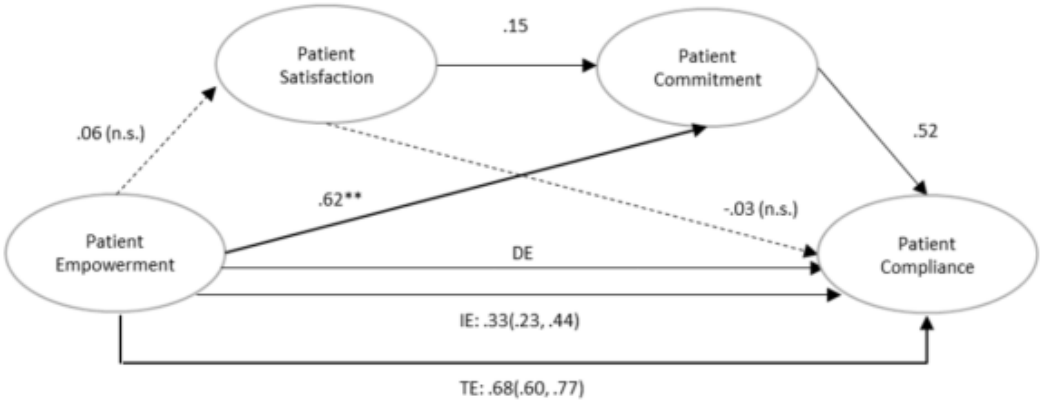
Estimated paths	Coefficient	SE	*t* test (*df*)	*P* value
Empowerment-Satisfaction	−0.01	0.09	−0.01 (35)	.76
Empowerment-Commitment	0.72	0.07	10.63 (35)	<.001
Satisfaction-Commitment	0.14	0.05	2.76 (35)	.005
Empowerment-Compliance	0.45	0.10	4.42 (35)	<.001
Satisfaction-Compliance	−0.05	0.04	−1.14 (35)	.05
Commitment-Compliance	0.54	0.10	5.28 (35)	<.001

Bootstrapping Model

A serial mediation analysis, also called multiple-step multiple mediation [45], using the bootstrapping SPSS PROCESS macro in SPSS (model 6) [46] on 10,000 resamples, gave a more precise estimate of the existence, polarity, and magnitude of the mediation effect of commitment and cross-validated the nonsignificance of satisfaction in the overall model. In the PROCESS macro in SPSS, the indirect mediation effect is considered to be significant when the confidence interval of the regression coefficient does not include 0 (Figure 3).

The bootstrapping results replicated the SEM findings; all the relationships were significant except that of empowerment on satisfaction (beta=.05; *P*=.33) and satisfaction on compliance (beta=−.03; *P*=.39). These findings are consistent with the SEM results in AMOS in SPSS. Also in line with the SEM procedure, empowerment affected commitment (beta=.62; *P*<.001) and satisfaction positively influenced commitment (beta=.15; *P*<.01), whereas commitment influenced compliance (beta=.51; *P*<.001). The direct effect of empowerment on compliance was significant (beta=.35; *P*<.001) as was the indirect effect (beta=.32; 95% CI 0.22-0.44), both making up for a highly significant total effect (beta=.68; *P*<.001). Although unrelated to empowerment, satisfaction slightly improved patient compliance indirectly through heightened commitment. These results suggest that commitment only partially explains the effect of empowerment on compliance [45]. In other words, it is through the direct effect of patient empowerment as well as the enhanced patient commitment triggered by patient empowerment that patient compliance grows.

Figure 3. Research model with bootstrapping with direct, indirect and total effects. DE: direct effect; IE: indirect effect; TE: total effect; ns: not significant.



Discussion

Principal Findings

This study investigated the direct and indirect influence of the empowerment perceived by patients of OHCs—specialized in chronic illnesses—on compliance with recommended treatment. We examined this effect via a mediational model linking empowerment to patient satisfaction and patient commitment in a multiple-step multiple mediation model.

This research relied on the use of a confirmatory approach that enables the simultaneous estimation and testing of several relationships among the predictive variables of empowerment, commitment, and satisfaction and their effects on compliance. To our knowledge, this is the first study to address the effects of patient empowerment gained through peer-to-peer OHCs on patient compliance with the proposed treatment. As such, this study extends prior research [14] showing that communication between physicians and patients in OHCs positively affects patient compliance. Lu and Zhang [14] found that when patients interact with physicians in OHCs, they have a better assessment of the quality of internet health information and they have better preferences during decision-making processes. Those results also underlined that patient-physician concordance is enhanced. These mechanisms, then in turn, positively impact patient compliance. In line with these findings, our research specific to peer-to-peer OHCs confirms that these communities can be used as a powerful tool to enhance patient compliance. This study also complements past research [13] showing that exchanging information with professional moderators in OHCs increases patient empowerment and improves cooperation

between the patient and the physician. It also extends previous research findings [24,25,47,48] highlighting that internet services contribute to enhance patient empowerment while demonstrating the underlying mechanism of how empowerment gained in OHCs enhances patient commitment toward the physician and patient compliance with the recommended treatment.

We used SEM not only because it allows for the testing of causal relationships among a set of variables within a hypothetico-deductive approach and because it estimates the strength of these relationships but also because it appraises measurement error. We also used the model 6 of the PROCESS macro in SPSS [45], on 10,000 resamples, to test the mediatory effects among the explanatory variables (Figure 3). Given the interrelationships that we hypothesized in our model (Figure 1), it was important to understand how such effects operate and whether some variables mediate the effect of other variables on the dependent variable, namely, patient compliance. Together, the SEM and Hayes PROCESS macro model approach underlined that 4 out of the 6 estimated paths of the research model were confirmed. Findings from this study highlight that out of the 3 antecedents of chronically ill patients' compliance that we model, 2 of them, namely, patient empowerment and commitment, had a direct and strong effect on the intended dependent variable of compliance. In other words, this study reveals that patient empowerment and patient commitment to the relationship with the physician have strong positive direct effects on chronically ill patients' compliance with recommended treatment, whereas patient satisfaction with the physician has no direct effect, but a mediating effect through patient commitment to the relationship with the physician, on compliance. Our data also highlight that patients' heightened sense of empowerment affects patient compliance not only directly but also indirectly by exerting a strong positive effect on patient commitment.

Prior research has shown that patient empowerment gained through peer-to-peer OHCs is determined by both the information utility found and shared in the communities and psychological benefits such as emotional support [49]. In the same vein, the computer-mediated social support gained through OHCs was found to positively alter the patient's commitment toward the physician [50]. Therefore, it is of strategic importance to posit both medical information sharing and social support as the core of the peer-to-peer OHC design. Specifically, patients should be encouraged to participate in peer-to-peer OHCs. As Johnston et al [49] underlined it, participation determines information utility and social support. Therefore, while designing OHCs communities, managers of those platforms should focus on implementing mechanisms that support the active participation of their members to elicit information sharing and enhance social support. Though still in development, the literature provides managers with useful guidelines about ways to stimulate such forms of participation and engagement. As participation in OHCs is intertwined with the perceived quality of the information shared [49], a *bottom-up* approach is advised to encourage self-regulation and self-rating processes of the information in these health support communities [51]. Third parties could evaluate the information shared, and there could be enforcement mechanisms in case of fraudulent or harmful information [51]. These mechanisms are expected to enhance the quality of the information shared and consequently patient's participation, which in turn will affect both the perceived information and perceived social support. Another avenue consists of providing tools to educate community users, to promote health literacy, and to help users feel confident to engage with those communities.

Interestingly, patient satisfaction with the physician did not appear to have a direct effect on patient compliance and was not determined by patient empowerment gained through OHCs (Figure 3).

Past research emphasized the emergence of dysfunctional empowerment emerging from OHCs, in that people with support from these communities may become less invested in their relationships with their physicians (eg, distrust of the physician, feelings of superiority over professional knowledge, and overconfidence in relation to the physician) [13]. Yet, our study shows that OHCs might also contribute to patients' compliance with prescribed treatment. The study underlines that satisfaction with the physician is not a significant contributor to the process, thus suggesting the occurrence of potential dysfunctional empowerment [13] that materializes in patient dissatisfaction. However, we did not control for this effect in the study, so it is difficult to estimate to what extent the low significance of satisfaction is related to dysfunctional empowerment. Importantly, the absence of influence of satisfaction does not prevent empowerment from exerting a significant effect on patients both directly and indirectly, suggesting that if dysfunctional empowerment is there, its effect is minimal in comparison with the overall positive influence of empowerment on compliance. This claim will need to be better substantiated by future research.

Practical Implications

Results from our study have several implications for health care services. First, they suggest that health care stakeholders should aim at enhancing patient empowerment on peer-to-peer OHCs. These communities provide their members with both informational and emotional support, enabling them to reduce uncertainty and make critical decisions about their health [48,52]. Defined as the ability to shape the composition of one's choice, patient empowerment positively influences patients' ability to make their own decisions about their health care [53]. Today, peer-to-peer OHCs contribute to the transfer of power and mastery from physicians to patients and favor improved health behaviors, better health outcomes, and reduced health care costs [11]. This study suggests that when empowered, peer-to-peer OHC members with a chronic disease report better compliance with their recommended treatments. It should be noted that this probably requires patients to have a reasonable level of literacy. Patients who do not have the technical health knowledge may not be able to join these communities. In the same vein, this suggests that patients are confident and competent enough to use peer-to-peer OHCs, which is not always the case. Therefore, efforts should be made to educate patients to overcome these shortcomings and avoid inadvertent exclusion of segment population. Knowing that half of all adults

worldwide have a chronic condition [3] and that 86% of internet users living with a chronic condition have searched online for medical information in 2007 [53], our findings are of considerable interest for both physician and health care systems. Our results indicate that peer-to-peer OHCs are powerful social tools that enhance chronically ill patients' commitment to the physician and compliance. This is why health care systems and physicians should promote peer-to-peer OHCs among patients suffering from a chronic condition.

Ultimately, this could be achieved by integrating these platforms in health care systems and physicians' workflows. However, research is needed to indicate how these communities could be integrated; we believe that patients could be invited to join these communities at an early stage of their relationship with the physician. This would probably require some time and effort from the physician or other health care workers to accompany the patient in his/her joining of these communities. This investment could prove to be beneficial in the mid- and long-run in the patient-physician relationship, bearing in mind that these communities also enhance disease self-care and reduce health care utilization [13]. While improving the workflow of physicians, these technological tools could contribute to a much broader phenomenon colloquially denominated as *shadow work* [54]. OHCs increase patients' workload of searching and analyzing vast quantities of information originating from patient exchange online, where previously the physician acted as the main filter and interpreter of medical information. The performance of this *shadow work* by the patient may decrease the physician's value in the eye of the patient. However, physicians remain central figures in the health care process, and with the continuous growth of these OHCs, doctors should act as information guides by helping patients to navigate through the complex net of information available to them online. This will make patients' information searches easier and more productive.

Overall, these results stress the importance of empowering patients and increasing commitment to the relationship that they have with their physician, and importantly, the possibility of using peer-to-peer OHCs as a means to do so. As peer-to-peer OHCs can help their members to gain more autonomy and efficacy in the self-management of their chronic disease, they positively influence users' well-being. As these communities reduce health care utilization, they contribute to a better use of health care resources. In that sense, our findings concur with those of Joglekar et al [55]. As OHCs provide the positive outcomes highlighted in this study and as they constitute a major source of health information for patients [56], health care stakeholders should encourage patients with chronic conditions to use peer-to-peer OHCs.

Limitations of This Study

Though this study highlights the implications of empowerment gained through peer-to-peer OHCs for compliance, we identified several limitations to the generalization of the results of this study. First, regarding the measurement of the empowerment construct, only 1 dimension, referring to physician support, was identified in our sample. It should also be noted that although we aimed at limiting the bias inherent to self-reported surveys, there remains potential bias, such as social desirability and selective recall [57]. Second, our model did not include all the predictors that have been reported in previous research so far. We focused on a key set of predictors, especially empowerment, because the literature has emphasized that this construct had a strong influence on patient compliance [58]. Prior research highlights that OHCs provide their members with the social support that leads them to feel a heightened sense of empowerment [59,60]. Other related constructs might be relevant as well, such as self-efficacy or perceived usefulness. Third, our SEM model conceptualized patient compliance and its predictors in a mediational analysis. However, some variables may moderate the relationships that we studied. In particular, we believe that patient literacy, and more specifically, electronic health (eHealth) literacy [61], moderates these relationships; it would be of interest to investigate its moderating effect on the paths we identified. Fourth, participants in our sample were French-speaking patients with chronic illness in the Province of Québec, Canada.

Directions of Future Research

To overcome the limitations mentioned in the previous section, future research should focus on the following specific aspects. First, further research should investigate the effect of the construct of social support gained through OHCs to increase our understanding of the effect of OHCs on chronically ill patients' compliance. Second, the moderating impact of patient literacy or eHealth literacy should be put to empirical test. Third, replication studies are needed to test the proposed theoretical framework in other provinces and countries with different health care systems. The external validity of the findings would increase if they can be replicated across various other medical care systems, contexts, and countries. Indeed, the Province of Québec, similar to Canada, offers a universal health care system to its citizens. While providing richer and more idiosyncratic insights into the impact of empowerment on compliance, it would be of interest to test our model in other types of health care systems that are entirely private or have a hybrid configuration. Fourth, future studies relying on qualitative methods, such as interviews or focus groups, would bring meaningful insights while limiting the bias inherent to self-reported studies.

Conclusions

The findings indicate that patient empowerment and commitment to the relationship with the physician are the 2 key predictors that enhance patient compliance with the prescribed treatment. Interestingly, patient satisfaction with the physician is found to impact patient compliance but through the mediating effect of patient commitment. Though patient empowerment has been shown to be critical in the health care literature, to the best of our knowledge, little research has empirically estimated how empowerment gained through peer-to-peer OHCs affected patients' propensity to comply with the prescribed treatment. This study suggests that health care stakeholders should encourage the use of peer-to-peer OHCs to enhance patient commitment and, ultimately, patient compliance with the physician. We believe that this research stream is and shall continue to be of considerable interest as compliance is a major

public health concern impacting the costs and performance of health care systems.

Acknowledgments

The authors thank the Fonds de développement académique du réseau (FODAR) from Université du Québec for their financial support. The authors also thank Dr K White, Professor of Psychology at Vanier College, Québec, Canada, and Dr K Gallopel-Morvan, Professor of Social Marketing at School of Public Health, France, for their helpful comments on the manuscript. This research was supported by the FODAR from Université du Québec. The views expressed are those of the authors.

Conflicts of Interest

None declared.

References

1. Lu X, Zhang R, Wu W, Shang X, Liu M. Relationship between internet health information and patient compliance based on trust: empirical study. J Med Internet Res 2018 Aug 17;20(8):e253 [FREE Full text] [doi: 10.2196/jmir.9364] [Medline: 30120087]
2. Tandon N, Mao J, Shprecher A, Anderson AJ, Cao F, Jiao X, et al. Compliance with clinical guidelines and adherence to antiretroviral therapy among patients living with HIV. Curr Med Res Opin 2018 Sep 15:1-9. [doi: 10.1080/03007995.2018.1519499] [Medline: 30173561]
3. Cutler RL, Fernandez-Llimos F, Frommer M, Benrimoj C, Garcia-Cardenas V. Economic impact of medication non-adherence by disease groups: a systematic review. BMJ Open 2018 Jan 21;8(1):e016982 [FREE Full text] [doi: 10.1136/bmjopen-2017-016982] [Medline: 29358417]
4. World Health Organization. Adherence to Long-term Therapies : Evidence for Action URL: http://www.who.int/iris/handle/10665/42682 [accessed 2018-12-29]
5. Spencer KL. Transforming patient compliance research in an era of biomedicalization. J Health Soc Behav 2018 Jun;59(2):170-184. [doi: 10.1177/0022146518756860] [Medline: 29443620]
6. Conthe P, Contreras EM, Pérez AA, García BB, Martín MN, Jurado MG, et al. Treatment compliance in chronic illness: current situation and future perspectives. Rev Clin Esp 2014;214(6):336-344. [doi: 10.1016/j.rce.2014.03.008] [Medline: 24816042]
7. Kleinsinger F. The unmet challenge of medication nonadherence. Perm J 2018;22:18-33 [FREE Full text] [doi: 10.7812/TPP/18-033] [Medline: 30005722]
8. Dolinski D, Dolinska B, Bar-Tal Y. Mood, cognitive structuring and medication adherence. Health Care Women Int 2018 Jul;39(7):808-820. [doi: 10.1080/07399332.2018.1458852] [Medline: 29608418]
9. Feehan M, Morrison MA, Tak C, Morisky DE, DeAngelis MM, Munger MA. Factors predicting self-reported medication low adherence in a large sample of adults in the US general population: a cross-sectional study. BMJ Open 2017 Jun 23;7(6):e014435 [FREE Full text] [doi: 10.1136/bmjopen-2016-014435] [Medline: 28645958]
10. Menvielle L, Audrain-Pontevia A, Menvielle W. The Digitization of Healthcare: New Challenges and Opportunities. London: Palgrave Macmillan; 2017.
11. Coleman WF. Cornerstone - Minnesota State University, Mankato. 2018. Optimism and Compliance: An Examination of Disempowering Processes Within Online Patient Communities URL: https://cornerstone.lib.mnsu.edu/eng_tech_comm_capstone_course/27 [accessed 2019-01-16]
12. Albright K, Walker T, Baird S, Eres L, Farnsworth T, Fier K, et al. Seeking and sharing: why the pulmonary fibrosis community engages the web 2.0 environment. BMC Pulm Med 2016 Jan 12;16:4 [FREE Full text] [doi: 10.1186/s12890-016-0167-7] [Medline: 26754048]
13. Petrič G, Atanasova S, Kamin T. Impact of social processes in online health communities on patient empowerment in relationship with the physician: emergence of functional and dysfunctional empowerment. J Med Internet Res 2017 Mar 13;19(3):e74 [FREE Full text] [doi: 10.2196/jmir.7002] [Medline: 28288953]
14. Lu X, Zhang R. Impact of physician-patient communication in online health communities on patient compliance: cross-sectional questionnaire study. J Med Internet Res 2019 May 13;21(5):e12891 [FREE Full text] [doi: 10.2196/12891] [Medline: 31094342]
15. Plaza V, Giner J, Calle M, Rytilä P, Campo C, Ribó P, et al. Impact of patient satisfaction with his or her inhaler on adherence and asthma control. Allergy Asthma Proc 2018 Nov 18;39(6):437-444 [FREE Full text] [doi: 10.2500/aap.2018.39.4183] [Medline: 30336791]
16. Prigge J, Dietz B, Homburg C, Hoyer WD, Burton JL. Patient empowerment: a cross-disease exploration of antecedents and consequences. Int J Res Mark 2015 Dec;32(4):375-386. [doi: 10.1016/j.ijresmar.2015.05.009]
17. Ouschan R, Sweeney J, Johnson L. Customer empowerment and relationship outcomes in healthcare consultations. Eur J Mark 2006 Sep;40(9/10):1068-1086. [doi: 10.1108/03090560610681014]

18. Chen L, Chen Y, Chen X, Shen X, Wang Q, Sun C. Longitudinal study of effectiveness of a patient-centered self-management empowerment intervention during predischarge planning on stroke survivors. Worldviews Evid Based Nurs 2018 Jun;15(3):197-205. [doi: 10.1111/wvn.12295] [Medline: 29878691]
19. Castro EM, van Regenmortel T, Vanhaecht K, Sermeus W, van Hecke A. Patient empowerment, patient participation and patient-centeredness in hospital care: a concept analysis based on a literature review. Patient Educ Couns 2016 Dec;99(12):1923-1939. [doi: 10.1016/j.pec.2016.07.026] [Medline: 27450481]
20. Bravo P, Edwards A, Barr PJ, Scholl I, Elwyn G, McAllister M, Cochrane Healthcare Quality Research Group, Cardiff University. Conceptualising patient empowerment: a mixed methods study. BMC Health Serv Res 2015 Jul 1;15:252 [FREE Full text] [doi: 10.1186/s12913-015-0907-z] [Medline: 26126998]
21. Rossi MC, Lucisano G, Funnell M, Pintaudi B, Bulotta A, Gentile S, BENCH-D Study Group. Interplay among patient empowerment and clinical and person-centered outcomes in type 2 diabetes. The BENCH-D study. Patient Educ Couns 2015 Sep;98(9):1142-1149. [doi: 10.1016/j.pec.2015.05.012] [Medline: 26049679]
22. Wilson TE, Kay ES, Turan B, Johnson MO, Kempf M, Turan JM, et al. Healthcare empowerment and HIV viral control: mediating roles of adherence and retention in care. Am J Prev Med 2018 Jun;54(6):756-764 [FREE Full text] [doi: 10.1016/j.amepre.2018.02.012] [Medline: 29656911]
23. Risling T, Martinez J, Young J, Thorp-Froslie N. Evaluating patient empowerment in association with eHealth technology: scoping review. J Med Internet Res 2017 Sep 29;19(9):e329 [FREE Full text] [doi: 10.2196/jmir.7809] [Medline: 28963090]
24. Samoocha D, Bruinvels DJ, Elbers NA, Anema JR, van der Beek AJ. Effectiveness of web-based interventions on patient empowerment: a systematic review and meta-analysis. J Med Internet Res 2010 Jun 24;12(2):e23 [FREE Full text] [doi: 10.2196/jmir.1286] [Medline: 20581001]
25. Riva S, Camerini A, Allam A, Schulz PJ. Interactive sections of an internet-based intervention increase empowerment of chronic back pain patients: randomized controlled trial. J Med Internet Res 2014 Aug 13;16(8):e180 [FREE Full text] [doi: 10.2196/jmir.3474] [Medline: 25119374]
26. Petrakaki D, Hilberg E, Waring J. Between empowerment and self-discipline: governing patients' conduct through technological self-care. Soc Sci Med 2018 Sep;213:146-153 [FREE Full text] [doi: 10.1016/j.socscimed.2018.07.043] [Medline: 30081356]
27. Mo PK, Coulson NS. Developing a model for online support group use, empowering processes and psychosocial outcomes for individuals living with HIV/AIDS. Psychol Health 2012;27(4):445-459. [doi: 10.1080/08870446.2011.592981] [Medline: 21854088]
28. Yeh M, Wu S, Tung T. The relation between patient education, patient empowerment and patient satisfaction: a cross-sectional-comparison study. Appl Nurs Res 2018 Feb;39:11-17. [doi: 10.1016/j.apnr.2017.10.008] [Medline: 29422144]
29. Audrain-Pontevia A, Menvielle L. Effects of interpersonal trust among users of online health communities on patient trust in and satisfaction with their physician. Int J Technol Assess Health Care 2018 Jan;34(1):56-62. [doi: 10.1017/S0266462317004433] [Medline: 29427998]
30. Cooper V, Clatworthy J, Harding R, Whetham J, EmERGE Consortium. Measuring empowerment among people living with HIV: a systematic review of available measures and their properties. AIDS Care 2019 Jul;31(7):798-802. [doi: 10.1080/09540121.2018.1537464] [Medline: 30428688]
31. Saiednejad Z, Ajorpaz NM, Aghajani M. Application of empowerment program on self-efficacy of patients under hemodialysis: a clinical randomized controlled trial. Crit Care Nurs J 2018 Sep 15;11(3):e81054. [doi: 10.5812/ccn.81054]
32. Náfrádi L, Nakamoto K, Schulz PJ. Is patient empowerment the key to promote adherence? A systematic review of the relationship between self-efficacy, health locus of control and medication adherence. PLoS One 2017;12(10):e0186458 [FREE Full text] [doi: 10.1371/journal.pone.0186458] [Medline: 29040335]
33. Bauer UE, Briss PA, Goodman RA, Bowman BA. Prevention of chronic disease in the 21st century: elimination of the leading preventable causes of premature death and disability in the USA. Lancet 2014 Jul 5;384(9937):45-52. [doi: 10.1016/S0140-6736(14)60648-6] [Medline: 24996589]
34. Hair JF, Black WC, Babin BJ, Anderson RE. Multivariate Data Analysis: Pearson New International Edition. Edinburgh, Scotland: Pearson Higher Education; 2014.
35. Oliver RL. A cognitive model of the antecedents and consequences of satisfaction decisions. J Mark Res 1980 Nov;17(4):460 [FREE Full text] [doi: 10.2307/3150499]
36. Morgan RM, Hunt SD. The commitment-trust theory of relationship marketing. J Mark 1994 Jul;58(3):20. [doi: 10.2307/1252308]
37. Schumacker RE, Lomax RG. A Beginner's Guide To Structural Equation Modeling: Third Edition. New York: Taylor & Francis Group; 2010.
38. Nunnally JC, Bernstein IH. Psychometric Theory. Third Edition. New York: Mcgraw-Hill; 1994.
39. Bulsara C, Styles I, Ward AM, Bulsara M. The psychometrics of developing the patient empowerment scale. J Psychosoc Oncol 2006;24(2):1-16. [doi: 10.1300/J077v24n02_01] [Medline: 17046803]
40. Meyers R, Pignault A, Houssemand C. Using the empowerment scale with unemployed people in lifelong learning: is the tool sound and useful? J Psy Res 2016 Nov 28;6(11):648-659. [doi: 10.17265/2159-5542/2016.11.003]
41. Rossiter JR. Reminder: a horse is a horse. Int J Res Mark 2005 Mar;22(1):23-25. [doi: 10.1016/j.ijresmar.2004.11.001]

42. Fornell CG, Larcker DF. Evaluating structural equation models with unobservable variables and measurement error. J Mark Res 1981 Feb;18(1):39. [doi: 10.2307/3151312]
43. Teo TS, Srivastava SC, Jiang L. Trust and electronic government success: an empirical study. J Man Info Syst 2014 Dec 8;25(3):99-132. [doi: 10.2753/MIS0742-1222250303]
44. Henseler J, Ringle CM, Sarstedt M. A new criterion for assessing discriminant validity in variance-based structural equation modeling. J Acad Mark Sci 2014 Aug 22;43(1):115-135. [doi: 10.1007/s11747-014-0403-8]
45. Hayes AF. Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach. New York: The Guilford Press; 2013.
46. Gutschoven K, van den Bulck J. Towards the Measurement of Psychological Health Empowerment in the General Public. In: Proceedings of the 56th Annual Conference of the International Communication Association. 2006 Presented at: ICA'06; June 19-23, 2006; Dresden, Germany.
47. Groen WG, Kuijpers W, Oldenburg HS, Wouters MW, Aaronson NK, van Harten WH. Empowerment of cancer survivors through information technology: an integrative review. J Med Internet Res 2015 Nov 27;17(11):e270 [FREE Full text] [doi: 10.2196/jmir.4818] [Medline: 26614438]
48. van Uden-Kraan CF, Drossaert CH, Taal E, Seydel ER, van de Laar MA. Self-reported differences in empowerment between lurkers and posters in online patient support groups. J Med Internet Res 2008 Jun 30;10(2):e18 [FREE Full text] [doi: 10.2196/jmir.992] [Medline: 18653442]
49. Johnston AC, Worrell JL, Di Gangi PM, Wasko M. Online health communities: an assessment of the influence of participation on patient empowerment outcomes. Info Technol People 2013 May 31;26(2):213-235. [doi: 10.1108/ITP-02-2013-0040]
50. Audrain-Pontevia A, Menvielle L. Do online health communities enhance patient-physician relationship? An assessment of the impact of social support and patient empowerment. Health Serv Manage Res 2018 Aug;31(3):154-162. [doi: 10.1177/0951484817748462] [Medline: 29280679]
51. Eysenbach G. Towards ethical guidelines for e-health: JMIR theme issue on eHealth ethics. J Med Internet Res 2000;2(1):E7 [FREE Full text] [doi: 10.2196/jmir.2.1.e7] [Medline: 11720926]
52. Brady E, Segar J, Sanders C. Accessing support and empowerment online: the experiences of individuals with diabetes. Health Expect 2017 Oct;20(5):1088-1095 [FREE Full text] [doi: 10.1111/hex.12552] [Medline: 28718928]
53. Fox S. Pew Internet Research. E-patients With a Disability or Chronic Disease URL: http://www.pewinternet.org/2007/10/08/e-patients-with-a-disability-or-chronic-disease/ [accessed 2019-01-15]
54. Lambert C. Shadow Work: The Unpaid, Unseen Jobs That Fill Your Day. Berkeley, California: Counterpoint; Apr 12, 2016.
55. Joglekar S, Sastry N, Coulson NS, Taylor SJ, Patel A, Duschinsky R, et al. How online communities of people with long-term conditions function and evolve: network analysis of the structure and dynamics of the asthma UK and British lung foundation online communities. J Med Internet Res 2018 Jul 11;20(7):e238 [FREE Full text] [doi: 10.2196/jmir.9952] [Medline: 29997105]
56. Wu B. Patient continued use of online health care communities: web mining of patient-doctor communication. J Med Internet Res 2018 Apr 16;20(4):e126 [FREE Full text] [doi: 10.2196/jmir.9127] [Medline: 29661747]
57. Althubaiti A. Information bias in health research: definition, pitfalls, and adjustment methods. J Multidiscip Healthc 2016;9:211-217 [FREE Full text] [doi: 10.2147/JMDH.S104807] [Medline: 27217764]
58. Umar A, Mundy D. Re-thinking models of patient empowerment. Stud Health Technol Inform 2015;209:175-181. [doi: 10.3233/978-1-61499-505-0-175] [Medline: 25980722]
59. Zhao T, Du R. Why People Are Willing to Provide Social Support in Online Health Communities: Evidence From Social Exchange Perspective. In: Proceedings of the International Conference on Smart Health. 2018 Oct 26 Presented at: ICSH'18; July 1-3, 2018; Wuhan, China p. 119-129.
60. Chen D, Zhang R, Liu K, Hou L. Enhancing online patient support through health-care knowledge in online health communities: a descriptive study. Information 2018 Aug 6;9(8):199. [doi: 10.3390/info9080199]
61. Levin-Zamir D, Bertschi I. Media health literacy, eHealth literacy, and the role of the social environment in context. Int J Environ Res Public Health 2018 Aug 3;15(8):pii: e1643 [FREE Full text] [doi: 10.3390/ijerph15081643] [Medline: 30081465]

Abbreviations

AVE: average variance extracted
CFI: comparative fit index
CMB: common method bias
eHealth: electronic health
FODAR: Fonds de développement académique du réseau
NFI: normed fit index
NNFI: nonnormed fit index
OHC: online health community
RMSEA: root mean square error of approximation

SEM: structural equation modeling
SRMR: standardized root mean residual

Edited by G Eysenbach; submitted 13.03.19; peer-reviewed by H Kondylakis, S Molinos, A Nguyen; comments to author 14.06.19; revised version received 06.08.19; accepted 31.08.19; published 11.11.19.

Please cite as:
Audrain-Pontevia AF, Menvielle L, Ertz M
Effects of Three Antecedents of Patient Compliance for Users of Peer-to-Peer Online Health Communities: Cross-Sectional Study
J Med Internet Res 2019;21(11):e14006
URL: https://www.jmir.org/2019/11/e14006
doi:10.2196/14006
PMID:31710295

Original Paper

Collective Empowerment in Online Health Communities: Scale Development and Empirical Validation

Sara Atanasova[1], PhD; Gregor Petric[1], PhD

Centre for Methodology and Informatics, Faculty of Social Sciences, University of Ljubljana, Ljubljana, Slovenia

Corresponding Author:
Sara Atanasova, PhD
Centre for Methodology and Informatics
Faculty of Social Sciences
University of Ljubljana
Kardeljeva pl 5
Ljubljana, 1000
Slovenia
Phone: 386 31 837 696
Email: sara.atanasova@fdv.uni-lj.si

Abstract

Background: The role of online health communities (OHCs) in patient empowerment is growing and has been increasingly studied in recent years. Research has focused primarily on individualistic conception of patients' empowerment, with much less attention paid to the role of OHCs in the development of patients' collective empowerment. Although OHCs have immense potential for empowerment that goes beyond the individual, the concept and scale of collective empowerment in OHCs have not yet been developed or validated.

Objective: This study aimed to develop an instrument for measuring collective empowerment in online health communities (CE-OHC) and to test its quality by investigating its factorial structure, reliability, construct validity, and predictive validity.

Methods: The CE-OHC scale was developed according to a strict methodology for developing valid and reliable scales. An initial set of 20 items was first tested in the pilot study conducted in 2016 using a sample of 280 registered users of Slovenia's largest OHC. A refined version with 11 items was tested in the main study conducted in 2018 on a random sample of 30,000 registered users of the same OHC. The final sample comprised 784 users. Exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) were used to investigate the factorial structure, discriminant validity, and convergent validity of the scale. Cronbach alpha coefficient was used to determine the CE-OHC scale's internal consistency. To establish the predictive validity, ordinary least squares regression was performed to test the role of CE-OHC in users' civic participation.

Results: The EFA resulted in a two-factor solution, and the two factors—knowledge of resources and resource mobilization for collective action—together explain 63.8% of the variance. The second-order CFA demonstrated a good fit to the data (root mean square error of approximation=0.07) and the scale had a good internal consistency (alpha=.86). Although evidence of the scale's convergent validity was partially provided, discriminant validity of the scale remained unconfirmed. Overall, CE-OHC was confirmed to be a predictor of users' civic participation, but the influence was somewhat weak and inconsistent across two subscales.

Conclusions: The proposed CE-OHC scale is a reliable and relatively valid instrument and serves as a good baseline to advance the measurement of collective empowerment in OHC contexts. This is the first scale developed for this purpose, and future research should focus on the development of a clear nomological network of the collective empowerment construct in relation to the OHC settings.

(J Med Internet Res 2019;21(11):e14392) doi:10.2196/14392

KEYWORDS

patient empowerment; collective empowerment; online health community; psychometrics; reliability; validity; weights and measures

Introduction

Online Health Communities as Platforms for Patient Empowerment

Online health communities (OHCs) are among the most important electronic health (eHealth) services in contemporary society [1]. OHC users, who are usually patients, caregivers, or other individuals interested in health-related issues, can search for and exchange health-related information, experiences, advice and social support, and/or influence public opinion and interact with other users and health professional moderators (usually doctors and health care providers), or simply observe others' interactions [1-9]. Several studies have demonstrated that the various activities of OHC users lead to patient empowerment [5,10,11], which is manifested in various positive outcomes for OHC users: higher self-esteem, self-efficacy, and control related to the management of one's health issues; enhanced satisfaction from helping others; improved confidence in interaction with doctors; more competent use of health services; and even enhanced social well-being and quality of life [6,12-15]. It is thus unsurprising that empowerment has become one of the central concepts within OHC studies. The importance of OHCs in patient empowerment has been increasingly acknowledged and studied in recent years; however, research has been primarily concerned with patients' individual empowerment, with little consideration of the role of OHCs in the development of patients' collective empowerment.

Lack of Research on Collective Empowerment

Patient empowerment is a predominately individualistic concept, originating from the general cultural shift in Western cultures toward individualism and consumerism in health care and focuses on various domains pertaining to patient, such as patient states and experiences, action and behaviors, self-determination, and skills [16]. Consequently, the research field is rich in measurement instruments that tap on some or all of the mentioned domains (refer to the study by Barr et al [16] for systematic review of such scales). Studies in health and health care, and OHCs specifically, focus almost exclusively on individual empowerment [13,17]. This is unsurprising as it has been demonstrated that if individuals experiencing health problems have positive attitudes, confidence, and other abilities required to manage their health, they may expect better health outcomes than individuals who are disengaged, apathetic, and resigned [18]. However, research of OHC suggests that at least in the context of this type of Web-based platforms, patients can also experience empowerment that is of intersubjective nature—called collective empowerment [6,15]. This finding is not specific to OHC research as the collective empowerment was already emphasized as important component of empowerment by the general empowerment theory from the field of community psychology. This theory clearly suggests that empowerment consists of at least 2 dimensions [19-22], that is, *individual* or *intrapersonal empowerment* and *collective empowerment* (also often referred to as *interactional* or *cognitive empowerment*) [23,24]. The domains of individual empowerment—such as abilities to develop a sense of control over personal health, self-efficacy, and competence in managing health conditions [20,25]—are clearly reflected in the concept of patient empowerment. Conversely, the ideas of developing psychological capacity for initiation or support of (potential) changes in social circumstances that affect patients' health conditions and the accessibility and quality of health services or health care system in general [20,26] are much less present in discussions of patient empowerment. This is considered by the concept of collective empowerment, which pertains to the individuals' beliefs that personal health-related issues can be (effectively) solved in collaboration with others and by enacting influence in wider social structures collectively [20,27].

The concept of collective empowerment is very relevant, at least in the context of OHCs, as these platforms allow users to engage in discussions of health politics and topics related to their lifestyles and values as individuals or groups. Moreover, social identity theory suggests that social identity processes in OHCs drive participatory behaviors and users' identification with the community, leading to users' collective engagement [28]. OHCs can thus function as communicative spaces in which, as in other types of online communities, participants may collectively engage and increase their social power as an interest group, with the aim of influencing the institutionalized arrangements and political decisions that affect their quality of life [29-31].

OHCs have indeed become an important arena in which individuals, patients, caregivers, and groups may voice their stances that challenge health policy, belief systems, practices in health care institutions, and services [32]. The bottom-up collective engagement that OHCs facilitate addresses topics that include access to or provision of health care services, health inequality, disease prevention and illness advocacy, health care reform, patients' rights, and power relationships in the health care arena. Moreover, OHCs often function as platforms for discussion and exchange of information related to the accessibility of remedies and medical treatments; access to health care services and health care professionals; misconceptions of specific, often stigmatized illnesses such as AIDS/HIV, infertility, and mental disorders; and other disease-related issues that often pertain to the disadvantaged social positions of specific patient groups [33,34].

Aim of the Study

The concept of collective empowerment has been studied in community psychology research [35,36], social identity theoretical perspectives [37,38] and, to some extent, implicitly investigated through concepts of patient engagement and activation in health studies [25,39]. However, studies of OHCs have investigated collective empowerment to a very limited extent [6,15,27,30,40-42] and, empirically, they offer no measurement instruments for assessing collective empowerment in OHCs. Hitherto, collective empowerment scales have been developed in community psychology research [36,43,44] and online community studies [6,30], but they have not been properly adapted to health-related contexts on the Web and empirically validated for the OHC setting. To develop a valid measure of collective empowerment in OHCs, the scale must consider the specific context of OHCs, which, compared with general online communities and other contexts, cover topics including personal health issues and discussions of health care

services and the health care system. As Zimmerman observed [20], the concept of empowerment is contextually dependent, varying across different populations and settings. Thus, it is crucial that any instrument used to measure collective empowerment is appropriately adapted to the relevant setting. Therefore, the aims of this study were (1) to develop a valid instrument for measuring collective empowerment in online health communities (CE-OHC) and (2) to test this instrument's quality by investigating its factorial structure, reliability, construct validity, and predictive validity.

Defining Collective Empowerment in Online Health Communities and Establishing Its Construct and Predictive Validity

Two Dimensions of Collective Empowerment

Collective empowerment has been generally defined as individuals' critical awareness and understanding of the sociopolitical environment [19,20] and thus consists of 2 main dimensions: (1) *knowledge of resources* and methods that can be used to impact social change and (2) *resource mobilization for collective action*.

Knowledge of resources refers to the application of individuals' knowledge and competences that might be used to collectively initiate change [15]. As already emphasized by early empowerment theorists [19,20], knowledge of resources comprises a critical assessment of individuals' social and political source(s) of their problem and the development of strategies aimed at collectively overcoming obstacles to achieving their goals. Web-based health-related settings such as OHCs play an important role in patients' processes of acquiring knowledge of the actions, strategies, or assets needed and applying this knowledge to address health-related problems. This knowledge may be acquired through user interactions whereby (collective) resources may be identified, potentially leading to collaborative efforts to develop strategies and solutions aimed at overcoming limitations in the issues affecting their health. For example, a qualitative study by Ammari and Schoenebeck [40] that explored online support groups for parents of children with special needs demonstrated that these parents were likely to connect, interact, and share their knowledge with other parents and to provide one another with insights into practices and strategies for addressing their child's health-related problems. Often, such collective efforts pushed parents to embrace advocacy beyond their own children's needs, leading to the development of interest and active individual and collective participation in legal, policy, and budgetary issues pertaining to their children's health conditions [40]. Without the knowledge of the resources required to resolve a specific problem that affects not only one individual but pertains more broadly to higher-order social structures, it is highly unlikely that individuals will be motivated to mobilize and influence the challenging social circumstances in a collective effort [45].

The second dimension of collective empowerment, *resource mobilization for collective action*, relates to individuals' awareness of the possibility for collective engagement and, with other individuals, the collective influence of arrangements in the specific social setting [19,45]. This dimension addresses individuals' recognition of the need for collaboration and coordination among larger groups—for example, community members—and for strengthening interpersonal relationships to exert an impact on the wider social circumstances that affect their lives and place them in a disadvantaged position [19]. OHCs have been shown to offer an important platform for the development of collective awareness and engagement that unite their members in a belief that personal health-related issues can be effectively solved through collaboration with others and by enacting collective influence in wider social structures [27,46].

Construct Validity

Construct validity is one of the most important indicators of a measurement instrument's quality as it pertains to the extent to which items in the scale actually measure what they are supposed to measure [47]. One common way to empirically assess construct validity is to investigate whether the proposed measure *behaves* as it should in relation to established measures of other constructs from the field. Empirically, construct validity is thus established via *convergent* and *discriminant validity*. The former refers to an empirical similarity between measures of theoretically related constructs and the latter pertains to the absence of correlation between measures of constructs that are theoretically unrelated [47].

On the basis of the theoretical and empirical evidence reported in previous studies, we identified 3 concepts that scholars emphasize to be correlated with collective empowerment (in OHCs) [19,27,30,41,48]: *sense of (virtual) community*, *involvement in community organization*, and *intensity of participation (in OHCs)*.

The crucial role that the sense of community plays in the development of collective empowerment was emphasized in early studies of empowerment in the field of community psychology [19,44,45]. In OHCs, a *sense of virtual community* is based on the users' identification with the online community, the perception of influence and emotional connection, and users' integration into the online community [49]. A sense of virtual community presents a key mechanism in building interpersonal relationships and developing awareness among online community members as it helps them to realize that their collaboration is essential for increasing social power as a group that can influence wider social structures [30]. This has also been emphasized by the social identity theory, which argues that identification with community as an important dimension of a sense of virtual community importantly leads to (effective) coordination, collaboration, collective action of members, and thereby, their collective empowerment [37]. The association between a sense of virtual community and collective empowerment in OHCs has been demonstrated by the studies by and [27,30]. These studies demonstrated that a sense of virtual community plays a crucial role in building collective empowerment in online communities as it helps users to develop responsibility for the community and a willingness to participate in supportive efforts. It also engenders a sense of social cohesion that encourages community members to collectively organize, develop a common goal, and engage in efforts to achieve it.

In addition, studies from the field of community psychology [19,45], as well as studies on online communities [30,48], attest to the importance of participation in a community's activities for the development of collective empowerment. *Involvement in community organization* pertains to online community members' inclusion in discussions about events, vision, and strategies of the online community [50,51]. Active engagement in community organization provides individuals with opportunities to learn new skills, interact with other members, identify needed resources, and develop critical awareness of one's environment [52].

There is also evidence that collective empowerment is associated with *different forms and intensities of participation in OHCs*. As the empowerment theory suggests, a certain investment in participation and active behavior is required to become empowered [20]. Thus, it is expected that users who contribute more to OHCs, post messages, and interact with other users (ie, posters) experience greater benefits and positive outcomes than users who participate passively in OHCs (ie, lurkers) or do not participate at all. The findings by  and  [6] and Li [41] indicate that participation plays an important role in collective empowerment as it has been confirmed that posters experience a higher level of collective empowerment than lurkers in OHCs.

To investigate discriminant validity, we examined the relationship between collective empowerment in OHCs and *received offline emotional support*. As collective empowerment in OHCs emerges with active participation in the Web-based platforms and the establishment of interpersonal relationships between users, there should be no association between the development of OHC users' collective empowerment and received offline emotional support. Users' development of collective empowerment in OHCs highly depends on internal (online) cohesiveness, group identification, and common goals that can lead users to seek political power together to advocate for social change [53], a process in which received offline emotional support should play no part. Moreover, received offline emotional social support leads to the fulfillment of sympathetic and caring behaviors, which is theoretically unrelated to the development of individuals' critical awareness, understanding of the sociopolitical environment, and knowledge of resources needed to initiate social change and collective action [54].

On the basis of the above theoretical and empirical underpinnings, we proposed the following hypothesis:

> *H1: Collective empowerment in OHCs is associated with a sense of virtual community, involvement in community organization, and intensity of participation in OHCs, but it is not associated with received offline emotional support.*

Predictive Validity

To establish predictive validity, the instrument used to measure a latent construct must be evaluated in terms of its ability to predict a certain form of behavior [55]. In our case, predictive validity pertained to the ability of the CE-OHC instrument to predict specific outcomes, such as individuals' actual participation in a wider sociopolitical environment and, thus, involvement in activities including petitions, demonstrations, and advocacy related to the issues of health-related public concern. In accordance with the empowerment theory [19,20], collective empowerment in OHCs should result in health-related *civic participation*.

Collective empowerment in OHCs refers to the development of a shared understanding among users of their position (as patients) in the wider social domain and leads to collective efforts in challenging existing health care institutions and services. An indirect link, at least, has been observed between collective empowerment and health-related civic participation in the case of an online support group for breast cancer patients in New Zealand. These patients, through participation in the online support group, identified an important issue regarding a national health insurance plan that did not cover a new treatment that, although expensive, was more efficient. With collective engagement and action in an online support group, these patients brought about a change in the national health insurance plan that introduced cover for new breast cancer treatments [33]. To establish the predictive validity of collective empowerment, other important factors of civic participation that have already been empirically validated must also be taken into account; these include a sense of (virtual) community [56], involvement in community organization [57], and intensity of participation in community settings [58]. On the basis of the above, we proposed the second hypothesis:

> *H2: Collective empowerment in OHCs is a significant predictor of civic participation.*

Methods

Sample

A pilot study and a main study were conducted to test the proposed CE-OHC scale. Both studies were based on data collected through a self-selected Web-based survey using probability samples of registered users of Slovenia's largest OHC, Med.Over.Net (MON). MON was founded in 2000 and covers areas of health, medicine, social work, law, and education. It is one of Slovenia's most visited online communities (and websites), with over 400,000 monthly visits and, on average, over 70,000 monthly users. In May 2018, this OHC had around 150 online discussion forums, around a million and a half forum threads, almost 12 million published forum posts, and around 120,000 registered users. The surveys for both studies were conducted using an open Web survey app 1KA (Eng. *One Click Survey*) developed by Centre for Social Informatics at Faculty of Social Sciences, University of Ljubljana. 1KA has mechanisms that disallow multiple entries by the same respondent. The data collection procedure and samples used in both studies are detailed in the following subsections.

Pilot Study

In preparation for the main study, a pilot study was conducted in June 2016 as part of the annual cross-sectional Web-based survey study on MON and its users. The Web survey was administered by the OHC provider and followed all ethical standards for the administration of scientific surveys. The Web

survey was conducted on a random sample of 15,000 registered users, and 13.04% (280/2147) of the respondents provided answers to the CE-OHC scale items. Details on the pilot study's data collection and sample may be found in the study by et al [59].

Main Study

The Web-based survey data collection for the main study was incorporated into an annual survey on users' experiences and satisfaction with the OHC MON, administered between April 25 and May 10, 2018, by the OHC provider and in line with the ethical standards for the administration of scientific surveys. The OHC provider designed a random sample of 30,000 registered users from the list of all registered users. Potential respondents were invited to participate in the Web survey via the OHC's email newsletter service. The invitation included a description of the study's purpose and brief information about respondents' rights and the survey length.

Out of approximately 30,000 potential respondents, 2314 (7.71%) clicked on the link for the Web survey, and 1762 respondents viewed the introduction page with informed consent and clicked the *Next* button to begin the survey. Of these, 676 (38.37%) partially completed and 893 (50.67%) fully completed the survey questionnaire, which led to a 76.15% (1762/2314) completion rate. The total response rate of 5.87% (1726/30,000) is small but not unusual in probability list–based Web surveys, which are long and include sensitive topics [60]. The survey questionnaire took on average 21 min and 33 seconds to complete. After the data screening and cleaning procedures, the final sample comprised 1123 respondents. The analyses were performed on a subsample of 784 respondents who had provided answers to the CE-OHC scale. Missing data were handled with the multiple imputation procedure. More information about the Web survey can be found in the Checklist for Reporting Results of Internet e-Surveys in Multimedia Appendix 1.

The sample consisted of 17.7% (139/784) males and 82.3% (645/784) females (Table 1). The respondents' average age was 41.1 years (SD 11.5). More than half of the participants of the study had completed higher education (64.7%, 507/784), were employed (72.2%, 566/784), and were married or de facto married (79.5%, 623/784). In the survey, the respondents were asked to self-asses how they perceived their current health on a scale of 1, poor, to 5, excellent. Most respondents reported having good (44.1%, 346/784) or very good (32.1%, 251/784) health status (Table 1).

Table 1. Sample characteristics (n=784).

Variable	Value, n (%)
Gender	
Male	139 (17.7)
Female	645 (82.3)
Education	
Lower	59 (7.5)
Middle	218 (27.8)
Higher	507 (64.7)
Employment status	
Employed or self-employed	566 (72.2)
Unemployed	75 (9.5)
School-aged youth or student	49 (6.3)
Retired	49 (6.3)
Homemaker or caregiver	34 (4.3)
Other	11 (1.4)
Marital status	
Married or de facto married	623 (79.5)
Single, divorced, or widowed	161 (20.5)
Health status	
Poor	8 (1.0)
Fair	72 (9.2)
Good	346 (44.1)
Very good	251 (32.1)
Excellent	107 (13.6)

Ethical Consideration

The authors of this study had no access to respondents' emails and received an anonymized dataset that included no identifiable personal information. No institutional ethics approval was required as this was a retrospective study. At all stages of the research process, we carefully protected all collected (personal) data and ensured participants' anonymity and confidentiality. The pilot and the main studies were also conducted in line with the Code of Ethics for Researchers at the University of Ljubljana [61] and the World Medical Association Declaration of Helsinki on ethical principles for medical research involving human subjects [62].

Measures

Collective Empowerment in Online Health Communities Scale

The scale's development followed the established process of operationalization from theoretical definition to development/adoption of items and empirical evaluation of the resulting scale [47,63]. By following a strict methodology for developing valid and reliable scales [47], and based on the identified dimensions of collective empowerment, we developed an initial set of 20 items. This item set was evaluated for content validity by 3 experts (1 in social science methodology, 1 in health communication, and 1 in internet studies), and a refined set of 15 items was selected, including 8 items measuring knowledge of resources and 7 items measuring resource mobilization for collective action. For knowledge of resources, 3 items were adopted from the study by Akey et al [64] and 5 were newly developed by the authors. The item set for *resource mobilization for collective action* comprised 2 items adopted from the study by Akey et al [64] and 7 items adopted from the Cognitive Empowerment Scale [19,45]. All items included in the final set for the CE-OHC scale were adapted and modified for the health-related and OHC contexts. The initial item set was pretested and evaluated in the pilot study. On the basis of the results of the pilot study (see Testing the Measurement Model: Pilot Study), we omitted 4 items, and the final version of CE-OHC scale included 11 items—6 (CE-OHC1 to CE-OHC6) for measuring *knowledge of resources* and 5 (CE-OHC7 to CE-OHC11) for measuring *resource mobilization for collective action*. All items were measured using a 5-point Likert-type scale, ranging from 1, strongly disagree, to 5, strongly agree.

Sense of Virtual Community

Sense of virtual community was measured by adapting 7 items from the Sense of Community Index [65,66] to the online community context. Respondents were asked to evaluate statements about the OHC forum that they most often visit on a scale of 1, strongly agree and 0, strongly disagree. The scale demonstrated acceptable internal consistency (alpha=.75).

Intensity of Participation in the Online Health Community

Intensity of participation in the OHC was measured with 9 items, which asked respondents to assess the frequency of their participation in online forum discussions (eg, posting, commenting, asking questions, opening new forum threads, and encouraging discussion) in the last 12 months on a 5-point scale of 1, never, to 5, very often. Responses to these items were summed in an index demonstrating good internal consistency (alpha=.91).

Involvement in Community Organization

Involvement in community organization was measured with 6 items. Respondents were asked about their engagement in activities relating to the OHC's vision, goals, and internal events. Answers were indicated on a 5-point scale of 1, never, to 5, very often. The 6 items were summed in an index that demonstrated good internal consistency (alpha=.95).

Civic Participation

Civic participation was measured with 4 items, asking respondents about their participation in activities related to initiatives and actions in the OHC that pertain to issues of public concern. Respondents indicated their answers on a 5-point scale of 1, never, to 5, very often. Civic participation items were summed in an index that demonstrated good internal consistency (alpha=.83).

Received Offline Emotional Support

To measure *received offline emotional support*, 3 items were used to ask respondents, on a 5-point scale of 1, never to 5, very often, how regularly they received various forms of emotional support from people in their everyday lives. This variable was computed as an aggregated average of its items and demonstrated good internal consistency (alpha=.92).

Control Variables

Membership length was measured by asking respondents to indicate how long they had been users of the OHC on the scale of 1, less than 1 month; 2, less than a year; 3, 1 to 3 years; and 4, more than 3 years.

Statistical Analyses

To test the measurement model of the CE-OHC scale and to ensure its construct validity [67], we first computed the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy and Bartlett test of sphericity (BTS) to determine whether our data were suitable for exploratory factor analysis (EFA). The KMO index ranges between 0 and 1, and values above 0.5 are considered suitable for factor analysis [67]. To ensure the suitability of factor analysis, BTS should be statistically significant ($P \leq .05$), which indicates that sufficient correlations exist among the variables [68]. EFA was conducted to determine which of the scale's items should be retained. Factors were extracted using principal axis factoring, which uses estimates of communalities on the diagonal in the extraction process [68]. As we did not expect an orthogonal factor solution, oblimin rotation was used.

Confirmatory factor analysis (CFA) was used to determine how well the measurement model fit the observed data [69]. As the construct of collective empowerment is composed of 2 latent dimensions, we used the second-order CFA approach to establish the construct validity of the CE-OHC scale. Second-order CFA is equivalent to *ordinary* first-order CFA, with the difference

that each latent dimension is modeled as an indicator of the second-order single latent construct [70]. In the case of 2 first-order factors, the model is underidentified [69]. To avoid the identification problem, the model requires additional information, which may be accomplished by including a constraint that sets the 2 factor loadings of the latent dimensions equal to one another [69]. To assess the model fit, the following absolute and incremental fit indices were used: (1) root mean square error of approximation (RMSEA, 0.08 as a cutoff for poor fitting models); (2) standardized root mean square residual (SRMR), where a value of less than 0.08 is generally considered a good fit; (3) comparative fit index (CFI), which ranges between 0.0 and 1.0, where values closer to 1.0 indicate good fit (CFI≥0.90); and (4) Tucker Lewis Index (TLI), which also ranges between 0.0 and 1.0, and where TLI≥0.9 indicates a good fit [71].

Construct validity was also assessed using a hypothesis-testing approach by which conceptual framework (theory) underlying the measure is used to state hypotheses regarding the relation between the measure and theoretically (un)related concepts and based on the empirical analysis and findings make inferences whether the measure is valid [72]. Correlation analysis with the Pearson correlation coefficient was used to test the association among CE-OHC, its subscales, and (un)related theoretical measures. The scale's reliability was assessed with the Cronbach alpha coefficient, which ranges between 0 and 1.0. The rough guidelines are that a value of .7 or higher indicated acceptable reliability and a value of .8 or better indicates good internal consistency [47].

To establish the scale's predictive validity, ordinary least squares (OLS) regression was performed to test the role of collective empowerment in OHCs in users' civic participation. Data were analyzed using IBM SPSS and R software, with the *lavaan* package [73] used for second-order CFA.

Results

Testing the Measurement Model: Pilot Study

Descriptive analysis of the CE-OHC items demonstrated that the majority are approximately normally distributed (see Multimedia Appendix 2). The interitem correlation analysis showed redundancy between 2 items pertaining to knowledge of resources (ie, "From using Med.Over.Net's forums, I know where to get information about resources needed to satisfy my health-related needs" and "From using Med.Over.Net's forums, I know how to get help from others to achieve my health-related goals"; r=0.82, *P*<.001) and 2 items relating to resource mobilization for collective action ("From using Med.Over.Net's forums, I have realized that the only way to improve health care in our country is by collaborating with other OHC users" and "From using Med.Over.Net's forums, I feel that I can only impact health care issues by working in an organized way with other OHC users"; r=0.82, *P*<.001). Consequently, we retained only one of the items for each of the above pairs (see Multimedia Appendix 2).

We conducted EFA to obtain communalities for each item and eigenvalues for extracted factors. The EFA showed a solution of 2 factors, whereby 2 items that were both reverse worded had communalities lower than 0.1 and a factor loading of around or below 0.2. Consequently, we omitted both items from further analysis. A principal axis factor analysis was conducted on 11 items of the CE-OHC scale (Multimedia Appendix 2).

The KMO measure was well above the acceptable limit of 0.5 [67] with KMO=0.89, which confirms the sampling adequacy of the analysis. The BTS was also statistically significant (*P*<.001), which indicates that sufficient correlations exist among the variables [68]. The EFA revealed that 2 factors had eigenvalues greater than Kaiser's criterion of 1 and, in combination, explained around 62.2% of the variance (Multimedia Appendix 2). All items on the CE-OHC scale had communalities around or above 0.4. Cronbach alpha, for both knowledge of resources and resource mobilization for collective action (alpha=.90), indicated good internal consistency. The obtained 2-factor solution was tested using CFA to verify the second-order measurement model of the CE-OHC scale (see Multimedia Appendix 2). The CFA fit indices suggested a reasonably good fit with χ^2_{43}=118.4, CFI=.96, TLI=.95, RMSEA=0.08, and SRMR=0.05. Factor loadings for both subscales were above 0.6, which supports the idea that CE-OHC is a single construct that manifests itself through the knowledge of resources and resource mobilization for collective action (Multimedia Appendix 2). Cronbach alpha (.90) for all items of the CE-OHC scale also supports this solution.

Testing the Measurement Model: Main Study

Of the 11 items of the CE-OHC scale, item CE-OHC2 had the highest mean (mean 3.72, SD 0.76) and item CE-OHC7 had the lowest mean (mean 2.41, SD 1.12; Table 2). A principal axis factor analysis was conducted on the 11 items using oblique rotation, direct oblimin (Table 2). The KMO measure confirmed the sampling adequacy for the analysis with KMO=0.86, which is well above the acceptable limit of 0.5 [67]. BTS was also statistically significant (*P*=.001), indicating data's suitability for factor analysis. An initial analysis was run to obtain eigenvalues for each factor in the data, with 2 factors having eigenvalues greater than Kaiser's criterion of 1 and, in combination, explaining 63.8% of the variance (Table 3). The communalities of all items were greater than 0.3. We retained 2 factors, and the items that cluster on the same factor suggest that the first factor represents *knowledge of resources* and the second factor represents *resource mobilization for collective action*. Cronbach alphas also indicate good internal consistency for both factors (alpha=.87 for knowledge of resources; alpha=.88 for resource mobilization for collective action; Table 3) and the correlation coefficient between factors is r=0.38.

Table 2. Factor loadings for collective empowerment in online health communities (CE-OHC) items and descriptive statistics (n=784).

Number of scale items	Scale items (From using Med.Over.Net's forums…)	Factor 1: Knowledge of resources	Factor 2: Resource mobilization for collective action	Value, mean (SD)
CE-OHC1	*...I know to whom I can turn when I have a health problem.*	0.70	0.02	3.56 (0.90)
CE-OHC2	*...I know how to use the health resources available to me in the OHC.*	0.79	−0.04	3.72 (0.76)
CE-OHC3	*...I know how to get help from others to achieve my health-related goals.*	0.82	−.02	3.55 (0.83)
CE-OHC4	*...I know how to access resources such as information, money, services, or support for dealing with health problems.*	0.74	−0.02	3.20 (0.98)
CE-OHC5	*...I understand better how our country's healthcare system works.*	0.58	0.07	3.13 (0.97)
CE-OHC6	*...I know which healthcare service I must use to solve my health problems.*	0.72	0.01	3.59 (0.84)
CE-OHC7	*...I actively advocate with other users for better healthcare in our country.*	−0.01	0.65	2.41 (1.12)
CE-OHC8	*...I feel that I can only impact healthcare issues by working in an organized way with other OHC users.*	−0.01	0.79	2.48 (1.11)
CE-OHC9	*...I believe that, to improve healthcare, it is more effective to work with a group of OHC users than as an individual.*	0.01	0.76	3.22 (1.12)
CE-OHC10	*...I realize that only by working together with other OHC users can we muster the power to change the healthcare system.*	0.01	0.87	2.90 (1.08)
CE-OHC11	*...I think that a user becomes powerful in the wider environment only through collaboration with other OHC users.*	0.01	0.79	3.09 (1.06)

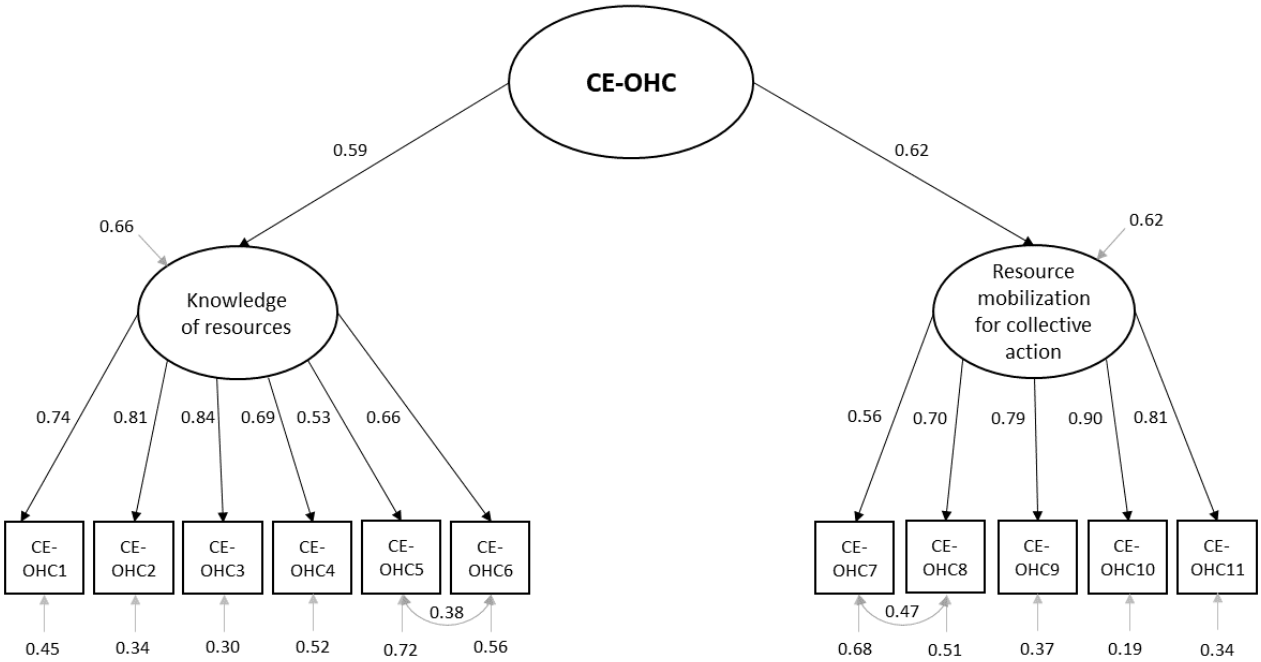
Table 3. Mean, standard deviation, percentage of variance, and Cronbach alphas of the two factors of collective empowerment in online health communities (CE-OHC) scale.

Factors of CE-OHC scale	Value, mean (SD)	Percentage of variance (%)	Alpha
Factor 1: Knowledge of resources	3.45 (0.68)	42.6	.87
Factor 2: Resource mobilization for collective action	2.82 (0.90)	21.2	.88

The 2-factor structure was tested using CFA to inspect the measurement model of the CE-OHC scale. On the basis of the assessment of the modification indices, we freed the covariance between 2 items of *knowledge of resources* (CE-OHC5 and CE-OHC6) and 2 items of *resource mobilization for collective action* (CE-OHC7 and CE-OHC8) as the items use similar phrasing. CFA revealed that the factor loadings for both factors are all above 0.5 (Figure 1). Fit indices with a revised parameter specification gave a better and a reasonably good fit (χ^2_{41}=208.9, CFI=.96, TLI=.95, RMSEA=0.07, and SRMR=0.04), which supports the CE-OHC scale's 2-dimensional structure (Figure 1). The CE-OHC scale's good internal consistency was also indicated by Cronbach alpha (.86).

Two subscales were created from the above items, where the knowledge of resources subscale has a higher mean (mean 3.45, SD 0.68) than does resource mobilization for collective action (mean 2.82, SD 0.90; Table 3). For further analysis, we also computed an overall *collective empowerment in OHCs* variable, which was calculated as the average of its subscales.

Figure 1. Second-order confirmatory factor analysis of collective empowerment in online health communities (CE-OHC) scale with standardized factor loadings of subscales and their items. CE-OHC1-11: Items of CE-OHC scale.



Discriminant and Convergent Validities

To verify the construct validity of the CE-OHC scale, correlation analysis was conducted to test the association among CE-OHC, its subscales, and suggested theoretically (un)related measures, that is, sense of virtual community, intensity of participation, involvement in community organization, and received offline emotional support (Table 4). As hypothesized with regard to convergent validity, the CE-OHC and its subscales, that is, knowledge of resources and resource mobilization for collective action, were significantly correlated with sense of virtual community and intensity of participation, although the correlation between knowledge of resources and intensity of participation is weak (r=0.11; *P*=.003). The results also demonstrated that CE-OHC and its subscale, resource mobilization for collective action, are weakly but significantly associated with involvement in community organization, whereas there is no significant association between the knowledge of resources subscale and involvement in community organization (r=0.02; *P*=.66).

Table 4. Bivariate correlations among collective empowerment in online health communities (CE-OHC) scale, its subscales, and theoretically (un)related measures (n=784).

CE-OHC scale and its subscales	Sense of virtual community		Intensity of participation		Involvement in community organization		Received offline emotional support	
	r	*P* value	*r*	*P* value	*r*	*P* value	*r*	*P* value
CE-OHC scale	0.44	<.001	0.27	<.001	0.12	.001	0.14	<.001
Knowledge of resources	0.30	<.001	0.11	.003	0.02	.66	0.18	<.001
Resource mobilization for collective action	0.40	<.001	0.30	<.001	0.18	.001	0.07	.06

With regard to the discriminant validity, we expected that the CE-OHC and its subscales would not be associated with received offline emotional support. The results, on the contrary, revealed a statistically significant and weak correlation between CE-OHC and received offline emotional support (*r*=0.14; *P*<.001) and between knowledge of resources and received offline emotional support (*r*=0.18; *P*<.001). As the results presented in Table 4 suggest, only the correlation between received offline emotional support and resource mobilization for collective action was not significant.

Predictive Validity

To assess the predictive validity of the CE-OHC scale and its subscales, we performed OLS regression. Two regression models were tested: one, in which the CE-OHC scale is included as an independent variable (*single model*), and the other, in which distinct subscales are included as independents variables (*subscale model*). We report the means with standard deviations of all (independent and dependent) variables included in the regression analysis in Table 5, and the results of both regressions are reported in Table 6.

Table 5. Descriptive statistics of variables in regression analysis (n=784).

Variables	Value, mean (SD)	Minimum	Maximum
Gender (0=male, 1=female)	0.82 (0.38)	0	1
Age (years)	41.1 (11.5)	18	90
Education	3.67 (0.80)	1	5
Membership length	3.69 (0.60)	1	4
Collective empowerment in online health communities (CE-OHC)	3.14 (0.65)	1	5
Knowledge of resources	3.45 (0.68)	1	5
Resource mobilization for collective action	2.82 (0.90)	1	5
Sense of virtual community	0.72 (0.27)	0	1
Intensity of participation	1.50 (0.62)	1	5
Involvement in community organization	1.15 (0.46)	1	4.33
Civic participation	1.24 (0.51)	1	4.75

Table 6. Multiple regression with civic participation as a dependent variable (n=784).

Predictor variables	Single model			Subscales model		
	b[a]	SE	beta	b	SE	beta
Collective empowerment in online health communities (CE-OHC)	0.10	0.02	.13[b]	—[c]	—	—
Knowledge of resources	—	—	—	0.03	0.02	.04
Resource mobilization for collective action	—	—	—	0.07	0.01	.12[b]
Sense of virtual community	−0.08	0.05	−.04[d]	-0.08	0.05	−.04[d]
Intensity of participation	0.13	0.02	.17[b]	0.13	0.02	.16[b]
Involvement in community organization	0.76	0.03	.69[b]	0.75	0.03	.69[b]
Membership length	0.01	0.02	.01	0.003	0.02	.01
Gender	0.02	0.04	.02	0.03	0.04	.02
Age	0.00	0.00	0.007	0.00	0.00	−.01
Education	0.01	0.02	.02	0.01	0.02	.02

[a]b: unstandardized regression coefficient.

[b]*P*<.001.

[c]The empty cells in the table are present because in the single model only collective empowerment in online health communities (CE-OHC) was included as independent variable and in the subscales model subscales knowledge of resources and resource mobilization for collective action were included as independent variables.

[d].05<*P*<.1.

In the single model, independent variables account for 66.4% of the variance in civic participation, and in the subscale model, independent variables account for 66.5% of the variance of the same dependent variable. The fit of each regression model was significant (F_{single}=191.5, *P*<.001; $F_{subscale}$=170.9, *P*<.001). As seen in Table 6, the overall CE-OHC scale demonstrated significant but very weak association with civic participation (beta=.13; *P*<.001). In the subscale model, only resource mobilization for collective action was significantly associated with civic participation (beta=.12; *P*<.001), although the knowledge of resources subscale was not significantly associated with the dependent variable (Table 6).

Among the predictors that were adopted from the literature, intensity of participation in OHCs (beta=.17; *P*<.001) and involvement in community organization (beta=.69; *P*<.001) were significantly associated with civic participation, whereas sense of virtual community was, according to the results (Table 6), very weakly and in the margins of statistical significance related to OHC users' civic participation. None of the control variables included in the regression models were statistically significantly associated with civic participation (Table 6).

Discussion

Principal Findings

The purpose of this study was to conceptualize collective empowerment in OHCs, to develop a scale to measure it, and to inspect the scale's psychometric properties in terms of

factorial structure, reliability, construct validity, and predictive validity.

First, we argued that collective empowerment is, in addition to individual empowerment, an important facet of the concept of patient empowerment with important ramifications for understanding the impact of OHCs on health care in general. Collective empowerment is composed of 2 distinct dimensions (*knowledge of resources* and *resource mobilization for collective action*), which was demonstrated by our pilot and main studies. Our study also showed that the CE-OHC scale may be considered a second-order factor, suggesting that the 11-item scale can be used to measure overall collective empowerment in OHC contexts. The measurement properties of the CE-OHC scale demonstrate that the scale is internally consistent with somewhat limited discriminant and predictive validities.

This study yielded an insight into the construct and predictive validity of the CE-OHC scale. To demonstrate the convergent and discriminant validities, we hypothesized that CE-OHC will be correlated with sense of virtual community, involvement in community organization, and intensity of users' participation in OHCs but not correlated with received offline emotional support. The results of the main study suggest that evidence for the CE-OHC scale's convergent validity can be only partially provided. Although the association between the CE-OHC scale (and its subscales) and sense of virtual community, as well as intensity of participation, was confirmed, the correlation between CE-OHC and involvement in community organization was very weak and even absent in the case of the knowledge of resources subscale. In other words, although users may be proactively involved in a community's activities and organization, it does not correlate with their collective empowerment, in terms of being able to understand the wider sociopolitical contexts of health issues and how to collectively engage to influence such contexts. We may speculate that such a result may be related to the fact that OHCs are complex entities that can include various subcommunities, for example, specific types of online discussion forums such as counseling or support group forums. Such subcommunities usually include specific types of community management, opportunity role structure, and sanctioning and monitoring mechanisms [74]. Involvement in community organization may thus rely heavily on the structure of a specific OHC's subcommunity, which, crucially, can also affect users' abilities to use and apply health-related knowledge acquired via the OHC, enabling them to resolve and address health issues within the health care system. In this study, involvement in community organization among OHC users was measured independently from the type of subcommunity or forum (online support group forum or online counseling forum) in which they most often participate. It is recommended that future research should focus on investigating how differences in the specific structural properties of OHC (subcommunities) affect users' patterns of involvement in online communities and collective empowerment.

Our study did not determine the CE-OHC scale's discriminant validity to be satisfactory. Offline emotional support is weakly associated with CE-OHC and its subscales, although such associations were not expected. Although research has not hitherto found evidence of a correlation between received offline emotional support and collective empowerment, there may exist a (in)direct link between the fulfillment of caring behaviors and emotional connection among people and their development of collective empowerment. Generally, developed support systems in (online) communities have been identified as an important facilitator of (individual) empowerment [75]. Further studies could focus on identifying the role of exchanged (offline) emotional and other types of social support in arriving at a critical understanding of the sociopolitical environment, knowledge of available resources, and methods of mobilizing those resources to collectively realize goals in the wider public domain. Moreover, future studies could also undertake a multitrait-multimethod approach [76] to assess discriminant and convergent validities and thus focus on investigating the patterns of the relationships between correlations of CE-OHC scale with similar and different constructs using different data collection methods.

In our study, the hypothesis regarding predictive validity was also only partially supported. An overall measure of the CE-OHC scale proved to be a significant predictor of civic participation, although the influence was somewhat weak. The subscale that pertains to knowledge of resources did not have a significant effect on users' engagement in activities pertaining to issues of public concern. Resource mobilization for the collective action subscale directly pertains to the perception of the collective power developed through an interaction with other individuals (users of the OHCs) and the possibilities of using this power to effect changes in the existing health care system. Thus, the significant direct effect is more plausible than it is in the knowledge of resources subscale, which pertains more to cognitive processes than to action. Although this subscale's validity is limited, we believe that both subscales present important components of the development of OHC users' collective empowerment. Resource mobilization is more activation oriented and predominately pertains to an awareness of the power that a collective effort, such as that of an OHC, has. However, resource mobilization also requires knowledge about certain issues, and awareness that these issues cannot be solved individually. Both components are then required for individuals, as members of the community, to gain influence as a whole and, consequently, generate change in the sociopolitical environment's structure. In the existing literature on collective empowerment, there is a lack of clear evidence of its specific outcomes or its antecedents, which are crucial for defining the nomological network of the collective empowerment construct. This scarcity of research is particularly evident in the health field and in OHC research, which has so far overlooked the importance of these Web-based platforms for the development of collective empowerment and, thus, its investigation. Consequently, the confirmation and validation of our results is currently beyond our reach. The CE-OHC scale will be an important baseline for future research into collective empowerment and should encourage research into this phenomenon and the processes surrounding it in various OHCs.

Limitations

This study has several limitations, making further research inevitable. In testing the convergent validity of the CE-OHC scale, we were limited by the aforementioned lack of a

nomological network that would have offered a stronger theoretical rationale for selecting criteria to establish validity. The evidence for discriminant validity was further hindered for the practical reason that we could not include additional variables in the already lengthy Web questionnaire. The scale's content validity could also be further improved by the inclusion of a larger set of experts to evaluate the initial pool of items and compute the content validity index.

Further attention should be also given to the optimization of some items' wording, as 2 pairs of items of each subscale appear to have common content that is disconnected from the content of the latent concept. However, we believe that the reliability of the overall CE-OHC scale, as well as of its subscales, was unaffected by this limitation as the results showed good internal consistency, which was confirmed in both the pilot and main studies.

Another limitation of this study is that both the pilot and main studies were geographically limited and conducted on a single OHC. MON is an OHC that has been in operation for almost 2 decades, and it has a high number and regular base of users and a structure that includes both online support group forums and counseling forums with specifically developed rules and norms that provide a foundation for the development of collective empowerment. Although Slovenia is typical of European Union countries with respect to the internet usage and information communication technologies [77], and MON is comparable with internationally established OHCs, such as PatientsLikeMe, MedHelp, and HealthUnlocked, the development and validation of the scale should include further studies covering different samples, cultures, and OHC settings.

Implications

This study carries several implications for research and practice. First, we are certain that the CE-OHC scale can be used for a plethora of quantitatively based investigations of emergent phenomena within new information communication technologies. Collective empowerment is an under-researched phenomenon, but it is key to understanding how individual activities transform into a psychological disposition for collective engagement and, in the next step, into actual collective action that has an impact on social change [43]. Explanatory and predictive research requires valid and reliable scales, and it is hoped that the proposal of the CE-OHC scale is a step toward the establishment of widely accepted standardized measures of collective empowerment. Moreover, as OHCs span across different Web-based platforms and include different stakeholders, the scale can be used to measure the collective empowerment of, for example, users of specific Facebook, Discord, Reddit, or Twitter groups, where units of interest may comprise patients, caregivers, and also nurses or doctors.

Second, measuring collective empowerment also carries implications for online community managers. Scales such as the CE-OHC scale allow community managers to assess the emerging social power within OHCs and the nature of this power and to implement measures aimed at managing such activities, for example, by providing functionalities that would help harness such power and providing mechanisms for generating an impact on wider social circumstances. In this context, measuring collective empowerment is also at least indirectly relevant for health policy makers in helping them to identify initiatives or even cases of patient unrest in Web-based platforms. We should note here that we assumed collective empowerment to be *functional*, in that it contributes to positive social change [5]. However, since empowerment can also be dysfunctional [5], this implies that collective empowerment can also be channeled toward unproductive or even damaging goals. Problematic collective behaviors can be evident, for example, in OHCs related to the antivaccination movement [78], proanorexia groups [79], or AIDS-denialist groups [34]. This implies that, when measuring collective empowerment, we should also consider measuring other variables pertaining to the goals of empowerment and other important predictors. Moreover, collective empowerment should be linked to eHealth literacy as previous research [5,59] has shown that a lack of eHealth literacy may lead to dysfunctional empowerment.

Conclusions

The 11-item CE-OHC scale appears to be a reliable and relatively valid instrument, developed to advance the measurement of collective empowerment in OHC contexts. To the best of our knowledge, the CE-OHC scale represents the first instrument developed for this purpose. The CE-OHC scale can help to identify the potential of OHCs to foster users' collective engagement and provide a framework that can inform the development of the resources needed to empower OHC social interactions. Research in collective empowerment generally transcends the individualistic approach and contributes to a more holistic understanding of empowerment as a process that brings together individuals, communities, and social change.

Acknowledgments

The authors acknowledge the help and technical support provided for the Web-based survey data collection by OHC Med.Over.Net and project manager A Verovšek. This research was supported by the Slovenian Research Agency (Research Project J5-9336).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Checklist for Reporting Results of Internet E-Surveys.
[PDF File (Adobe PDF File), 280 KB - jmir_v21i11e14392_app1.pdf]

Multimedia Appendix 2
Pilot study results.
[PDF File (Adobe PDF File), 416 KB - jmir_v21i11e14392_app2.pdf]

References

1. Johnston AC, Worrell JL, di Gangi PM, Wasko M. Online health communities: an assessment of the influence of participation on patient empowerment outcomes. Info Technol People 2013;26(2):213-235. [doi: 10.1108/ITP-02-2013-0040]
2. Coulson NS, Bullock E, Rodham K. Exploring the therapeutic affordances of self-harm online support communities: an online survey of members. JMIR Ment Health 2017 Oct 13;4(4):e44 [FREE Full text] [doi: 10.2196/mental.8084] [Medline: 29030324]
3. Huh J, Kwon BC, Kim S, Lee S, Choo J, Kim J, et al. Personas in online health communities. J Biomed Inform 2016 Oct;63:212-225 [FREE Full text] [doi: 10.1016/j.jbi.2016.08.019] [Medline: 27568913]
4. Peng X, Sun D, Zhao Y, Xu W. What Trigger People Use Physician-patient Interactive OHCs? An Empirical Research Based Integration Model. In: Proceedings of the Pacific Asia Conference on Information Systems. 2015. Presented at: PACIS'15; July 6-9, 2015; Singapore URL: https://aisel.aisnet.org/pacis2015/2/
5. Petric G, Atanasova S, Kamin T. Impact of social processes in online health communities on patient empowerment in relationship with the physician: emergence of functional and dysfunctional empowerment. J Med Internet Res 2017 Mar 13;19(3):e74 [FREE Full text] [doi: 10.2196/jmir.7002] [Medline: 28288953]
6. Petrovcic A, Petric G. Differences in intrapersonal and interactional empowerment between lurkers and posters in health-related online support communities. Comput Human Behav 2014;34:39-48. [doi: 10.1016/j.chb.2014.01.008]
7. van der Eijk M, Faber MJ, Aarts JW, Kremer JA, Munneke M, Bloem BR. Using online health communities to deliver patient-centered care to people with chronic conditions. J Med Internet Res 2013 Jun 25;15(6):e115 [FREE Full text] [doi: 10.2196/jmir.2476] [Medline: 23803284]
8. Vennik FD, Adams SA, Faber MJ, Putters K. Expert and experiential knowledge in the same place: patients' experiences with online communities connecting patients and health professionals. Patient Educ Couns 2014 May;95(2):265-270. [doi: 10.1016/j.pec.2014.02.003] [Medline: 24598314]
9. Yang H, Guo X, Wu T. Exploring the influence of the online physician service delivery process on patient satisfaction. Decis Support Syst 2015 Oct;78:113-121 [FREE Full text] [doi: 10.1016/j.dss.2015.05.006]
10. Mo PK, Coulson NS. Empowering processes in online support groups among people living with HIV/AIDS: a comparative analysis of 'lurkers' and 'posters'. Comput Human Behav 2010;26(5):1183-1193. [doi: 10.1016/j.chb.2010.03.028]
11. Visser LM, Bleijenbergh IL, Benschop YW, Van Riel AC, Bloem BR. Do online communities change power processes in healthcare? Using case studies to examine the use of online health communities by patients with Parkinson's disease. BMJ Open 2016 Nov 7;6(11):e012110 [FREE Full text] [doi: 10.1136/bmjopen-2016-012110] [Medline: 27821596]
12. Bartlett YK, Coulson NS. An investigation into the empowerment effects of using online support groups and how this affects health professional/patient communication. Patient Educ Couns 2011 Apr;83(1):113-119. [doi: 10.1016/j.pec.2010.05.029] [Medline: 20599338]
13. Mo PK, Coulson NS. Developing a model for online support group use, empowering processes and psychosocial outcomes for individuals living with HIV/AIDS. Psychol Health 2012;27(4):445-459. [doi: 10.1080/08870446.2011.592981] [Medline: 21854088]
14. van Uden-Kraan CF, Drossaert CH, Taal E, Seydel ER, van de Laar MA. Participation in online patient support groups endorses patients' empowerment. Patient Educ Couns 2009 Jan;74(1):61-69. [doi: 10.1016/j.pec.2008.07.044] [Medline: 18778909]
15. Wentzer HS, Bygholm A. Narratives of empowerment and compliance: studies of communication in online patient support groups. Int J Med Inform 2013 Dec;82(12):e386-e394. [doi: 10.1016/j.ijmedinf.2013.01.008] [Medline: 23523082]
16. Barr PJ, Scholl I, Bravo P, Faber MJ, Elwyn G, McAllister M. Assessment of patient empowerment--a systematic review of measures. PLoS One 2015;10(5):e0126553 [FREE Full text] [doi: 10.1371/journal.pone.0126553] [Medline: 25970618]
17. Anderson RM, Funnell MM. Patient empowerment: myths and misconceptions. Patient Educ Couns 2010 Jun;79(3):277-282 [FREE Full text] [doi: 10.1016/j.pec.2009.07.025] [Medline: 19682830]
18. Schulz PJ, Nakamoto K. Health literacy and patient empowerment in health communication: the importance of separating conjoined twins. Patient Educ Couns 2013 Jan;90(1):4-11. [doi: 10.1016/j.pec.2012.09.006] [Medline: 23063359]
19. Speer PW. Intrapersonal and interactional empowerment: implications for theory. J Community Psychol 2000;28(1):51-61. [doi: 10.1002/(sici)1520-6629(200001)28:1<51::aid-jcop6>3.3.co;2-y]
20. Zimmerman MA. Psychological empowerment: issues and illustrations. Am J Community Psychol 1995 Oct;23(5):581-599. [doi: 10.1007/BF02506983] [Medline: 8851341]
21. Zimmerman MA. Taking aim on empowerment research: On the distinction between individual and psychological conception. Am J Community Psychol 1990;18(1):169-177. [doi: 10.1007/bf00922695]
22. Hur MH. Empowerment in terms of theoretical perspectives: exploring a typology of the process and components across disciplines. J Community Psychol 2006 Sep;34(5):523-540. [doi: 10.1002/jcop.20113]

23. Peterson NA, Hamme CL, Speer PW. Cognitive empowerment of African Americans and Caucasians. J Black Stud 2002;32(3):336-351 [FREE Full text] [doi: 10.1177/002193470203200304]
24. Wilke LA, Speer PW. The mediating influence of organizational characteristics in the relationship between organizational type and relational power: an extension of psychological empowerment research. J Community Psychol 2011;39(8):972-986. [doi: 10.1002/jcop.20484]
25. Palumbo R. The Bright Side and the Dark Side of Patient Empowerment: Co-creation and Co-destruction of Value in the Healthcare Environment. Cham, Switzerland: SpringerBriefs in Public Health; 2017.
26. Israel BA, Checkoway B, Schulz A, Zimmerman MA. Health education and community empowerment: conceptualizing and measuring perceptions of individual, organizational, and community control. Health Educ Q 1994;21(2):149-170. [doi: 10.1177/109019819402100203] [Medline: 8021145]
27. Petrovcic A, Petric G. Factors of collective psychological empowerment of active users in the online health community med.over.net. Slovenian J Public Heal 2014;53(2):133-143. [doi: 10.2478/sjph-2014-0014]
28. Liu N, Tong Y, Chan HC. Information seeking in online healthcare communities: the dual influence from social self and personal self. IEEE Trans Eng Manage 2017;64(4):529-538. [doi: 10.1109/tem.2017.2703121]
29. Heidelberger CA, El-Gayar O, Sarnikar S. Online Health Social Networks and Patient Health Decision Behavior: A Research Agenda. In: Proceedings of the 44th Hawaii International Conference on System Sciences. 2011 Presented at: HICSS'11; Jan 4-7, 2011; Kauai, HI, USA. [doi: 10.1109/hicss.2011.328]
30. Petric G, Petrovcic A. Individual and collective empowerment in online communities: the mediating role of communicative interaction in web forums. Inform Soc 2014;30(3):184-199. [doi: 10.1080/01972243.2014.896683]
31. Swan M. Emerging patient-driven health care models: an examination of health social networks, consumer personalized medicine and quantified self-tracking. Int J Environ Res Public Health 2009 Feb;6(2):492-525 [FREE Full text] [doi: 10.3390/ijerph6020492] [Medline: 19440396]
32. Orr D, Baram-Tsabari A, Landsman K. Social media as a platform for health-related public debates and discussions: the Polio vaccine on Facebook. Isr J Health Policy Res 2016;5:34 [FREE Full text] [doi: 10.1186/s13584-016-0093-4] [Medline: 27843544]
33. Radin P. 'To me, it's my life': medical communication, trust, and activism in cyberspace. Soc Sci Med 2006 Feb;62(3):591-601. [doi: 10.1016/j.socscimed.2005.06.022] [Medline: 16039031]
34. Rykov YG, Meylakhs PA, Sinyavskaya YE. Network structure of an AIDS-Denialist online community: identifying core members and the risk group. American Behavioral Scientist 2017;61(7):688-706. [doi: 10.1177/0002764217717565]
35. Speer PW. Social power and forms of change: implications for psychopolitical validity. J Community Psychol 2008 Mar;36(2):199-213. [doi: 10.1002/jcop.20231]
36. Peterson NA, Lowe JB, Aquilino ML, Schneider JE. Linking social cohesion and gender to intrapersonal and interactional empowerment: support and new implications for theory. J Community Psychol 2005 Mar;33(2):233-244 [FREE Full text] [doi: 10.1002/jcop.20047]
37. Drury J, Reicher S. Collective psychological empowerment as a model of social change: researching crowds and power. J Soc Iss 2009;65(4):707-725 [FREE Full text] [doi: 10.1111/j.1540-4560.2009.01622.x]
38. Drury J, Evripidou A, van Zomeren M. Empowerment: The intersection of identity and power in collective action. In: Sindic D, Barreto M, Costa-Lopes R, editors. Power and Identity. London: Psychology Press; 2014:102-124.
39. Mazzoni D, Alberici AI. Why do patients protest? Collective action processes in people with chronic illnesses: A psychosocial perspective. In: Khosrow-Pour M, editor. Health Care Delivery and Clinical Science: Concepts, Methodologies, Tools, and Applications. Hershey, PA, USA: IGI Global; 2018:1197-1219.
40. Ammari T, Schoenebeck S. Networked Empowerment on Facebook Groups for Parents of Children with Special Needs. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. 2015 Presented at: CHI'15; April 18-23, 2015; Seoul, Republic of Korea. [doi: 10.1145/2702123.2702324]
41. Li Z. Psychological empowerment on social media: Who are the empowered users? Public Relat Rev 2016;42(1):49-59 [FREE Full text] [doi: 10.1016/j.pubrev.2015.09.001]
42. Lundström R. Social support online: between closeness and anonimity. In: Lindgren S, editor. Hybrid Media Culture: Sensing Place in the World of Flows. London & New York: Routledge; 2014:106-122.
43. Christens BD, Collura JJ, Tahir F. Critical hopefulness: a person-centered analysis of the intersection of cognitive and emotional empowerment. Am J Community Psychol 2013 Sep;52(1-2):170-184. [doi: 10.1007/s10464-013-9586-2] [Medline: 23793530]
44. Speer PW, Peterson NA, Armstead TL, Allen CT. The influence of participation, gender and organizational sense of community on psychological empowerment: the moderating effects of income. Am J Community Psychol 2013 Mar;51(1-2):103-113. [doi: 10.1007/s10464-012-9547-1] [Medline: 22847224]
45. Speer PW, Peterson NA. Psychometric properties of an empowerment scale: Testing cognitive, emotional, and behavioral domains. Soc Work Res 2000 Jun 1;24(2):109-118. [doi: 10.1093/swr/24.2.109]
46. van Uden-Kraan CF, Drossaert CH, Taal E, Shaw BR, Seydel ER, van de Laar MA. Empowering processes and outcomes of participation in online support groups for patients with breast cancer, arthritis, or fibromyalgia. Qual Health Res 2008 Mar;18(3):405-417. [doi: 10.1177/1049732307313429] [Medline: 18235163]

47. DeVellis RF. Scale Development: Theory And Applications. Thousand Oaks: Sage Publications; 2012.
48. Petric G, Petrovcic A. Elements of the management of norms and their effects on the sense of virtual community. Online Inform Rev 2014;38(3):436-454 [FREE Full text] [doi: 10.1108/oir-04-2013-0083]
49. Blanchard AL. Developing a sense of virtual community measure. Cyberpsychol Behav 2007 Dec;10(6):827-830. [doi: 10.1089/cpb.2007.9946] [Medline: 18085972]
50. Leung L. User-generated content on the internet: an examination of gratifications, civic engagement and psychological empowerment. New Media Soc 2009 Nov 24;11(8):1327-1347. [doi: 10.1177/1461444809341264]
51. Novek EM. Communication and community empowerment. Peace Review 1999;11(1):61-68 [FREE Full text] [doi: 10.1080/10402659908426231]
52. Zimmerman MA. Empowerment theory: psychological, organizational and community levels of analysis. In: Rappaport J, Seidman E, editors. Handbook of Community Psychology. New York: Kluwer Academic; 2000:43-63.
53. Mok BH. Organizing self-help groups for empowerment and social change: Findings and insights from an empirical study in Hong Kong. J Community Pract 2005;13(1):49-67 [FREE Full text] [doi: 10.1300/j125v13n01_04]
54. Speer PW, Jackson CB, Peterson NA. The relationship between social cohesion and empowerment: support and new implications for theory. Health Educ Behav 2001 Dec;28(6):716-732. [doi: 10.1177/109019810102800605] [Medline: 11720274]
55. Allen MJ, Yen WM. Introduction to Measurement Theory. Monterey, CA, USA: Waveland Press, Inc; 1979.
56. Cho H, Jahng J. Factors affecting the performance of voluntary participants in the knowledge sharing virtual community. Int J Web Based Communities 2014;10(3):339-356 [FREE Full text] [doi: 10.1504/IJWBC.2014.062946]
57. Perkins DD, Brown BB, Taylor RB. The ecology of empowerment: predicting participation in community organizations. J Soc Iss 1996;52:85-110. [doi: 10.1111/j.1540-4560.1996.tb01363.x]
58. Ognyanova K, Chen NT, Ball-Rokeach SJ, An Z, Son M, Parks M, et al. Online participation in a community context: civic engagement and connections to local communication resources. Int J Commun 2013;7:2433-2456 [FREE Full text]
59. Petric G, Atanasova S, Kamin T. Ill literates or illiterates? Investigating the eHealth literacy of users of online health communities. J Med Internet Res 2017 Oct 4;19(10):e331 [FREE Full text] [doi: 10.2196/jmir.7372] [Medline: 28978496]
60. Callegaro M, Manfreda KL, Vehovar V. Web Survey Methodology. Los Angeles: Sage Publications, Ltd; 2015.
61. University of Ljubljana. 2009. Code of Ethics. URL: https://www.uni-lj.si/university/code_of_ethics/ [accessed 2016-11-15] [WebCite Cache ID 6m2ZtYJmq]
62. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. J Am Med Assoc 2013 Nov;310(20):2191-2194. [doi: 10.1001/jama.2013.281053] [Medline: 24141714]
63. Hayes BE. How to measure empowerment. Quality Progress 1994;27(2):41-46 [FREE Full text]
64. Akey TM, Marquis JG, Ross ME. Validation of scores on the psychological empowerment scale: A measure of empowerment for parents of children with a disability. Educ Psychol Meas 2000;60(3):419-438. [doi: 10.1177/00131640021970637]
65. Chavis DM, Hogge JH, McMillan DW, Wandersman A. Sense of community through Brunswik's lens: a first look. J Community Psychol 1986;14(1):24-40 [FREE Full text] [doi: 10.1002/1520-6629(198601)14:1<24::aid-jcop2290140104>3.0.co;2-p]
66. McMillan DW, Chavis DM. Sense of community: a definition and theory. J Community Psychol 1986 Jan;14(1):6-23 [FREE Full text] [doi: 10.1002/1520-6629(198601)14:1<6::aid-jcop2290140103>3.0.co;2-i]
67. Hair JF, Anderson RE, Tatham RL, Black WC. Multivariate Data Analysis. New Jersey: Prentice-Hall Inc; 1998.
68. Meyers LS, Gamst G, Guarino AJ. Applied Multivariate Research: Design And Interpretation. Los Angeles: Sage Publications Inc; 2016.
69. Kline RB. Principles and Practice of Structural Equation Modeling. New York & London: Guilford; 1998.
70. Bollen KA. Structural Equations With Latent Variables. New York: John Wiley & Sons; 1989.
71. Hooper D, Coughlan J, Mullen M. Structural equation modelling: guidelines for determining model fit. Elecron J Bus Res Methods 2008;6(6):53-60 [FREE Full text]
72. Waltz C, Strickland O, Lenz E. Measurement in Nursing and Health Research. Fourth Edition. New York: Springer Publishing Company; 2010.
73. Rosseel Y. Lavaan: an R package for structural equation modeling and more. Version 0.5-12 (BETA). J Stat Soft 2012;48(2):1-36. [doi: 10.1002/9781119579038.ch1]
74. Atanasova S. Research Gate. 2018. Socio-structural Properties and Psychological Empowerment in Online Health-related Support Communities. URL: https://www.researchgate.net/publication/329753856_Socio-structural_properties_and_psychological_empowerment_in_online_health-related_support_communities [accessed 2019-09-15] [WebCite Cache ID 77eDtmL7E]
75. Maton KI, Salem DA. Organizational characteristics of empowering community settings: a multiple case study approach. Am J Community Psychol 1995 Oct;23(5):631-656. [doi: 10.1007/BF02506985] [Medline: 8851343]
76. Campbell DT, Fiske DW. Convergent and discriminant validation by the multitrait-multimethod matrix. Psychol Bull 1959 Mar;56(2):81-105. [doi: 10.4135/9781412961288.n79] [Medline: 13634291]

77. Statistical Office of the Republic of Slovenia. 2017. URL: https://pxweb.stat.si/pxweb/Dialog/varval.asp?ma=2974205S&ti=&path=../Database/Ekonomsko/23_29_informacijska_druzba/11_IKT_posamezniki/04_29742_uporaba_inter/&lang=2 [accessed 2019-09-15] [WebCite Cache ID 77eErEdCH]
78. Kata A. Anti-vaccine activists, Web 2.0, and the postmodern paradigm--an overview of tactics and tropes used online by the anti-vaccination movement. Vaccine 2012 May 28;30(25):3778-3789. [doi: 10.1016/j.vaccine.2011.11.112] [Medline: 22172504]
79. Gavin J, Rodham K, Poyer H. The presentation of 'pro-anorexia' in online group interactions. Qual Health Res 2008 Mar;18(3):325-333. [doi: 10.1177/1049732307311640] [Medline: 18235156]

Abbreviations

BTS: Bartlett test of sphericity
CE-OHC: collective empowerment in online health communities
CFA: confirmatory factor analysis
CFI: comparative fit index
EFA: exploratory factor analysis
eHealth: electronic health
KMO: Kaiser-Meyer-Olkin
MON: Med.Over.Net
OHC: online health community
OLS: ordinary least squares
RMSEA: root mean square error of approximation
SRMR: standardized root mean square residual
TLI: Tucker Lewis Index

Edited by G Eysenbach; submitted 15.04.19; peer-reviewed by T Risling, J Mcgeough; comments to author 03.07.19; revised version received 05.09.19; accepted 05.09.19; published 20.11.19.

Please cite as:
Atanasova S, Petric G
Collective Empowerment in Online Health Communities: Scale Development and Empirical Validation
J Med Internet Res 2019;21(11):e14392
URL: http://www.jmir.org/2019/11/e14392/
doi:10.2196/14392
PMID:31746772

Review

Reducing Potentially Inappropriate Prescriptions for Older Patients Using Computerized Decision Support Tools: Systematic Review

Luís Monteiro[1,2], MD; Tiago Maricoto[3,4], MD; Isabel Solha[5], MD; Inês Ribeiro-Vaz[2,6,7], PhD; Carlos Martins[2,7], PhD; Matilde Monteiro-Soares[2,7], PhD

[1]Esgueira+ Family Health Unit, Aveiro Healthcare Centre, Aveiro, Portugal

[2]Center for Health Technology and Services Research, Faculty of Medicine, University of Porto, Porto, Portugal

[3]Aveiro-Aradas Family Health Unit, Aveiro Healthcare Centre, Aveiro, Portugal

[4]Faculty of Health Sciences, University of Beira Interior, Covilhã, Portugal

[5]Terras de Souza Family Health Unit, Paredes, Portugal

[6]Porto Pharmacovigilance Centre, Faculty of Medicine, University of Porto, Porto, Portugal

[7]Department of Community Medicine, Information and Decision in Health, Faculty of Medicine, University of Porto, Porto, Portugal

Corresponding Author:
Luís Monteiro, MD
Esgueira+ Family Health Unit, Aveiro Healthcare Centre
Rua Pedro Vaz de Eça
Aveiro
Portugal
Phone: 351 00351234312890
Email: luismonteiro.net@gmail.com

Abstract

Background: Older adults are more vulnerable to polypharmacy and prescriptions of potentially inappropriate medications. There are several ways to address polypharmacy to prevent its occurrence. We focused on computerized decision support tools.

Objective: The available literature was reviewed to understand whether computerized decision support tools reduce potentially inappropriate prescriptions or potentially inappropriate medications in older adult patients and affect health outcomes.

Methods: Our systematic review was conducted by searching the literature in the MEDLINE, CENTRAL, EMBASE, and Web of Science databases for interventional studies published through February 2018 to assess the impact of computerized decision support tools on potentially inappropriate medications and potentially inappropriate prescriptions in people aged 65 years and older.

Results: A total of 3756 articles were identified, and 16 were included. More than half (n=10) of the studies were randomized controlled trials, one was a crossover study, and five were pre-post intervention studies. A total of 266,562 participants were included; of those, 233,144 participants were included and assessed in randomized controlled trials. Intervention designs had several different features. Computerized decision support tools consistently reduced the number of potentially inappropriate prescriptions started and mean number of potentially inappropriate prescriptions per patient. Computerized decision support tools also increased potentially inappropriate prescriptions discontinuation and drug appropriateness. However, in several studies, statistical significance was not achieved. A meta-analysis was not possible due to the significant heterogeneity among the systems used and the definitions of outcomes.

Conclusions: Computerized decision support tools may reduce potentially inappropriate prescriptions and potentially inappropriate medications. More randomized controlled trials assessing the impact of computerized decision support tools that could be used both in primary and secondary health care are needed to evaluate the use of medication targets defined by the Beers or STOPP (Screening Tool of Older People's Prescriptions) criteria, adverse drug reactions, quality of life measurements, patient satisfaction, and professional satisfaction with a reasonable follow-up, which could clarify the clinical usefulness of these tools.

Trial Registration: International Prospective Register of Systematic Reviews (PROSPERO) CRD42017067021; https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42017067021

(J Med Internet Res 2019;21(11):e15385) doi:10.2196/15385

KEYWORDS

deprescriptions; medical informatics applications; potentially inappropriate prescription; potentially inappropriate medication; computerized decision support

Introduction

The older adult population is increasing in developed countries [1], and people worldwide are living longer [2,3]. According to the World Health Organization, people aged 60 years and older in 2020 will outnumber children younger than 5 years. In 2050, the world's population aged 60 years and older is expected to total 2 billion [2].

The aging of populations increases the pressure on health care systems, which should be aligned with the needs of older populations [4]. Older patients are more likely to have more than one chronic condition, known as multimorbidity [5,6]. The prevalence of multimorbidity is more than 90% in older patients [5]. Having more than one chronic condition requires the use of several medications. Thus, older adults are more vulnerable to polypharmacy [7], meaning the use of multiple drugs administered to the same patient [8,9], in addition to prescriptions of potentially inappropriate medications (PIMs) [10-12]. A PIM can be described as a medication use that has potentially more risks than benefits with a safer alternative available [10].

Potentially inappropriate prescription (PIP) is a broader concept than PIM, because it includes over-, under-, and misprescribing (eg, inappropriate dose or duration). It is defined as "the prescribing of medication that could introduce a significant risk of an adverse event, in particular when there is an equally or more effective alternative with lower risk available" [13].

Due to changes in pharmacokinetics and pharmacodynamics, older people are more prone to drug interactions and adverse drug reactions [14,15]. Adverse drug reactions are considered a public health problem in older patients and a cause of disability and mortality [15]. Deprescribing is defined as "the process of withdrawal of inappropriate medication, supervised by a health care professional, with the goal of managing polypharmacy and improving outcomes" [16].

There are several ways to address polypharmacy to prevent its occurrence [17-23]. This review focused on computerized decision support (CDS) tools. Bates et al [24] defined CDS systems as computer-based systems providing "passive and active referential information as well as reminders, alerts, and guidelines." Payne [25] added that CDS tools can be defined as "computer applications designed to aid clinicians in making diagnostic and therapeutic decisions in patient care." CDS tools may have a positive impact on health care, such as reducing physicians' orders of unnecessary tests [26].

Previous studies reviewed such strategies, such as multidisciplinary team medication reviews, pharmacist medication reviews, computerized clinical decision support systems, and multifaceted approaches and reported substantial heterogeneity in the included studies, but did not focus on CDS [19,21]. One systematic review that did focus on CDS systems included studies published only through 2012, and new studies have been published since then [27]. This systematic review aims to clarify whether CDS tools can help in reducing PIPs or PIMs to improve clinical outcomes in older adults.

Methods

Eligibility Criteria

The systematic review was conducted according to a protocol previously published [28] and registered in PROSPERO (International Prospective Register of Systematic Reviews; CRD42017067021). We searched for interventional controlled studies (type of study) with participants aged 65 years or older (population) that assessed whether CDS tools (intervention) could diminish PIM (outcome). Moribund or terminal participants were excluded along with those requiring palliative care. No other restriction was applied.

Search Methods

We searched MEDLINE, CENTRAL, EMBASE, and Web of Science for studies published through February 2018 without language restrictions. Specific queries were used according to each database's requirements that were described in detail elsewhere [29]. Trial registries, different types of grey literature, and contact with specialists in the field were also performed. The reference lists of all included studies were searched to identify any potentially pertinent study that might not have been identified by previous methods. References were checked from previously published systematic reviews.

Selection Process

Articles were selected by applying the criteria to the title and abstract of each study. Studies that were selected at this stage were then assessed in their entirety. Each stage was conducted by two researchers blindly and independently. Two reviewers (LM and TM) examined the titles and abstracts and did the full-text screening. When disagreement occurred, it was resolved through consensus.

Data Collection Process

For all the included studies, characterization of data and results were exported into a datasheet by one of the authors (LM) and confirmed by the other (MS).

Type of Data Collected

Studies were characterized according to setting, intervention, comparison definition, study duration, number of included participants overall and in each study group, the proportion of missing data, participants' mean age, the proportion of male individuals, and deprescribing target. Outcomes retrieved from each study were categorized as PIP- or PIM-related and by overall number of prescriptions, adverse drug reactions, and potential drug-drug interactions.

Analysis of Results and Assessment of the Risk of Bias

Possible bias in randomized controlled trials (RCTs) was independently identified using the Cochrane Collaboration Risk

of Bias tool [29] by two researchers (TM and LM). This assessment was confirmed by other authors (IV and MS). Risk of bias was determined with regard to random sequence generation, allocation concealment, blinding of participants and personnel, blinding of outcome assessments, incomplete outcome data, selective outcome reporting, and other biases.

The included articles did not permit the performance of a meta-analysis because there were not a minimum of three studies using the same deprescribing target. Thus, only a narrative synthesis was performed. We have summarized the main features and results of all the included studies, discussed their limitations, and proposed future research avenues.

Results

Description of the Studies

Using our search strategy, 3756 articles were identified through MEDLINE, Central, EMBASE, and Web of Science databases. One article was identified through contact with specialists. After duplicates were removed, 2819 articles remained. The titles and abstracts were screened, and 2767 studies were excluded. Of these, 52 articles were selected to assess eligibility and their full text was analyzed. Of these, 36 articles were excluded. Ultimately, we included 16 studies in our systematic review. No new article was found by searching in the included studies' reference lists, trial registries, or grey literature. The article selection process and reasons for exclusion are described in Figure 1.

The characteristics of the included studies are described in Table 1. More than half (10/16) of the included studies were RCTs, one was a crossover study, and five were pre-post intervention studies. Most studies were conducted in North America (Canada and United States; n=11) [30-40]. The remaining were conducted in Europe (n=5) [41-45].

Six studies were conducted exclusively in secondary health care institutions [35,37,38,40,44,45]. In two studies, only emergency department participants were included [33,39]. In total, six studies were performed exclusively in primary health care institutions [30-32,41-43], one study took place in a health maintenance organization [34], and one study included participants from both secondary and primary health care institutions [36]. Six studies took place at teaching hospitals [36-38,40,44,45].

Most commonly, the standard of care was the only comparator (n=11). The interventional design was always based on a CDS tool, which was usually included in the electronic medical record with several different features. In some cases (n=6), complex interventions were performed that included training and engagement sessions and/or leaflet provision.

The RCTs had an inclusion period ranging from 3 to 30 months (see Table 2). The crossover study included four on-off periods with a 6-week duration [33]. The pre-post intervention studies frequently compared different time periods.

Figure 1. Flow diagram on search and article inclusion, according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Statement.

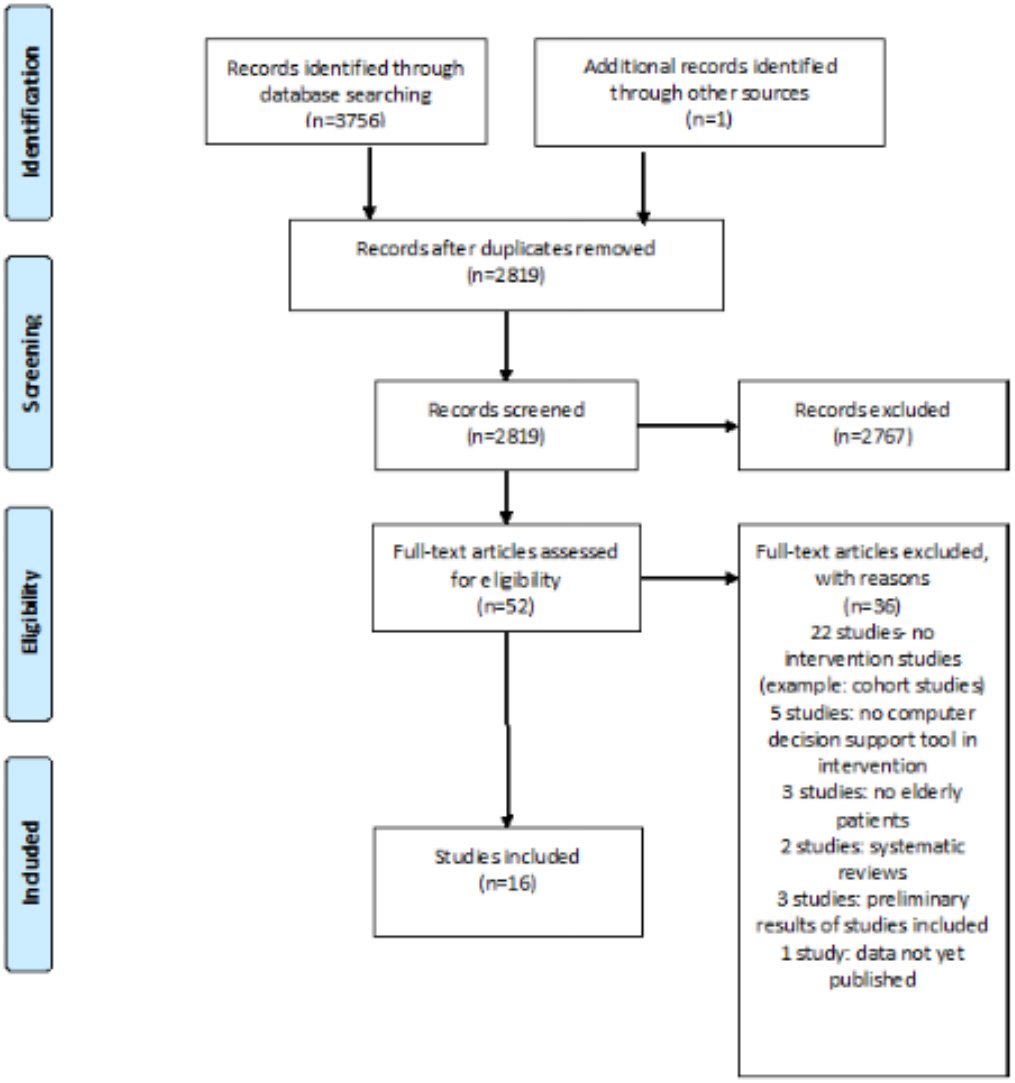


Table 1. Descriptions of the included studies in the systematic review (N=16).

Author, year; (study); country	Setting	Comparator	Intervention	Deprescribing target
Randomized controlled trials				
Tamblyn et al [30], 2003; Canada	PHC[a]	Usual care[b]	Computerized decision support tool providing alert identified problem + presented possible consequences + provided alternative therapy	PIP[c] (159 clinically relevant PIPs in the elderly defined by expert consensus)
Price et al [31], 2017; Canada	PHC (8 GP[d])	Usual care	Clinical decision support tool showing alert with specific STOPP[e] guideline content in electronic medical record	PIPs (40 STOPP criteria)
Avery et al [41], 2012; (PINCER); UK	PHC (72 GP)	Computer-generated simple feedback	PINCER; comparator + pharmacist-led information technology complex intervention	PIPs on NSAIDs[f], beta blockers, ACE[g] inhibitors, or loop diuretics
Erler et al [42], 2012; Germany	PHC (46 GP)	Usual care	Interactive 1-hour workshop for physicians on detection and management of CKD[h] + provision of desktop checklist of medications to be reduced or avoided + patient information leaflets + training in the use of software "DOSING"	Prescription exceeding recommended standard; daily dosage >30% or recommended; maximum daily dose in CKD patients
Clyne et al [43], 2015; (OPTI-SCRIPT); Ireland	PHC (21 GP)	Usual care + simple, patient-level PIP postal feedback	Comparator + academic detailing with pharmacist + medicine review with Web-based pharmaceutical treatment algorithms + leaflets	PIPs using 28 criteria from the study
Cossette et al [40], 2017; Canada	SHC[i] (teaching hospital)	Usual care	KT[j] strategy; distribution of educational materials + in-services by geriatricians + computerized alert systems pharmacist-physician	7 PIMs[k] based Beers and STOPP geriatric criteria and drugs with anticholinergic properties or acting on the central nervous system
Fried et al [32], 2017; (TRIM); USA	PHC (Veterans Affairs; medical center)	Usual care only and usual care with telephonic patient assessment	2 Web apps: (1) extracts information on medications and chronic conditions from the electronic health record, (2) interface for data chart review and telephonic patient assessment + a set of automated algorithms evaluating medication appropriateness + patient-specific medication management feedback report for the clinician	Medication appropriateness based on range of criteria, including feasibility in context of patient's cognition and social support, potential overtreatment of DM[l] or hypertension, "traditional" PIMs according to Beers and STOPP criteria, inappropriate renal dosing, and patient report of adverse medication effects
O'Sullivan et al [44], 2016; Ireland	SHC (teaching hospital)	Usual medical and pharmaceutical care	Clinical decision support software supported structured pharmacist review of medication designed to optimize geriatric pharmaceutical care	Medicines associated with "nontrivial" adverse drug reactions (according to WHO)
Terrel et al [33], 2009; USA	ED[m] (teaching hospital)	Computerized; physician order entry without alerts	Computer-assisted decision support alert when PIM was being prescribed + rationale + recommended safer substitute therapies. If physician chose to continue, second menu displayed to query most important reason	9 high-use and high-impact PIMs[n]
Raebel et al [34], 2007; USA	HMO[o] (18 medical offices + 21 pharmacies)	Usual care	Medication alert generated from PIMS not allowing prescription label to be printed until the pharmacist actively determined whether prescription should be dispensed; pharmacists should communicate notifications to prescribing clinicians	Newly prescribed PIMs based on the Beers, Zhan and Kaiser Performance Care Management Institute lists of medications to be avoided in older people[p]
Crossover studies				

Author, year; (study); country	Setting	Comparator	Intervention	Deprescribing target
Peterson et al [35], 2005; USA	SHC	Usual computerized order entry	Guided dosing of psychotropic medication integrated in Brigham Integrated Computer System	Benzodiazepines, opiates, and neuroleptics
Pre-post intervention studies				
Ruhland et al [36], 2017; USA	SHC + PHC; (1 teaching hospital + 2 community hospital + 31 clinics)	Usual care	Clinical decision support system creating an alert + rational and; alternative medication through Epic (an integrated electronic medical record)	PIMs on glyburide
Mattinson et al [37], 2010; USA	SHC (teaching hospital)	Usual care	Medication-specific warning system (advised alternative medication or dose reduction)	PIMs on medications not recommended for use in older patients (not recommended medications) and those for which only a reduced dose was advised (dose-reduction medications)
Lester et al [38], 2015; USA	SHC (teaching hospital)	Computerized physician order entry without alerts	Computerized; physician order entry with pop-up alerts for selected PIPs containing links to articles relevant to the alert	PIPs on diphenhydramine, metoclopramide, and antipsychotics
Ghibelli et al [45], 2013; (INTERcheck); Italy	SHC (teaching hospital)	Analysis without any interference	Computer-based application (INTERCheck) that collects, stores and automatically; provides drug information to reduce or prevent PIPs	PIMs from 2003 Beers Criteria; potential DDIs[q]; and Anticholinergic Cognitive Burden Scale
Stevens et al [39], 2017; (EQUiPPED); USA	ED (10 Veterans Affairs; medical centers)	Usual care	EQUiPPED interventions: education + informatics-based clinical decision support + individual provider feedback	PIMs from 2012 Beers Criteria category 1 (to avoid in all older adults)

[a]PHC: primary health care.

[b]Each physician was given a computer, printer, health record software, and access to the internet.

[c]PIP: potentially inappropriate prescription.

[d]GP: general practice.

[e]STOPP: Screening Tool of Older People's Prescriptions.

[f]NSAID: nonsteroidal anti-inflammatory drug.

[g]ACE: angiotensin-converting enzyme.

[h]CKD: chronic kidney disease.

[i]SHC: secondary health care.

[j]KT: knowledge translation.

[k]PIM: potentially inappropriate medication.

[l]DM: diabetes mellitus.

[m]ED: emergency department.

[n]High-use and high-impact PIMs: promethazine, diphenhydramine, diazepam, propoxyphene with acetaminophen, hydroxyzine, amitriptyline, cyclobenzaprine, clonidine, indomethacin.

[o]HMO: health maintenance organization.

[p]Examples of medications to be avoided in older people: amitriptyline, chlordiazepoxide, chlorpropamide, diazepam, doxepin, flurazepam, aspirin in combination with hydrocodone or oxycodone, ketorolac, oral meperidine, and piroxicam.

[q]DDI: drug-drug interaction.

Table 2. Characterization of the included studies in the systematic review, including study type, study duration, sample size, and participant demographics (N=16).

Study	Study duration (months); date range	Sample size, N	Participants, n			Outcome missing data, n (%)
				Age (years), mean (SD)	Gender (male), n (%)	
Randomized controlled trials						
Tamblyn et al [30]	13; (01/1997-02/1998)	12,560	C[a]: 6276; I[b]: 6284	C: 75 (6); I: 75 (6)	C: 2248 (36); I: 2439 (39)	N/R[c]
Price et al [31]	8; (02-10/2015)	81,905	C:37,615; I: 44,290	N/R; all >65 years	N/R	N/R
Avery et al [41]	6 (and 12)	480,942	C: 37,659; I: 34,413	N/R	N/R	C: 22 (0.06); I: 28 (0.08) for outcome 3
Erler et al [42]	6	404	C: 206; I: 198	C: 80 (9); I: 81 (6)	C: 63 (31); I: 81 (41)	C: 9 (4); I: 0 (0)
Clyne et al [43]	6; (10/2012-09/2013)	196	C: 97; I: 99	C: 76 (5); I: 77 (5)	C: 50 (52); I: 55 (56)	C: 3 (3); I: 3 (3)
Cossette et al [40]	10 weeks; (09/2015-12/2015)	321	C: 133; I: 139	C: 81 (7); I: 82 (8)	C: 53 (41); I:48 (38)	C: 5 (4); I: 13 (9)
Fried et al [32]	3; (10/2014-01/2016)	156	C1: 36; C2: 39; I: 81	<70 years C: 25 (39); I: 27 (42)	C: 63 (99); I: 63 (99)	C1: 4 (11); C2:7 (18); I: 17 (21)
O'Sullivan et al [44]	13; (06/2011-07/2012)	737	C: 361; I: 376	C: 78b; (IQR 72-84); I: 77; (IQR 71-83)	C: 190 (51); I: 180 (50)	C: 17 (5); I: 17 (5)
Terrel et al [33]	30; (12/01/2005-07/07/2007)	5162	C: 2515; I: 2647	C: 74 (7); I: 74 (7)	C: 880 (35); I: 929 (35)	N/R
Raebel et al [34]	12; (18/05/2005-17/05/2006)	59,680	C: 29,840; I: 29,840	C: 74; (5-95 percentile 66-88); I: (5-95 percentile 66-88)	C: 12,843 (43); I: 12704 (43)	N/R
Crossover studies						
Peterson et al [35]	4 × 6 week on-off periods; (08/10/2001-16/05/2002)	3718	C: 1925; I: 1793	C: 75 (7); I: 75 (7)	C: 905 (47); I: 843 (47)	N/R
Pre-post intervention studies						
Ruhland et al [36]	3 + 3; (B[d]: 01/12/2014-28/02/2015); A[e]: 01/03/2015-31/05/2015)	N/R	101 patients with activated alert	75	N/R	N/A[f]
Mattison et al [37]	6 + 41.5; (B: 1/06-29/11/2014; A: 17/03/2015-30/08/2008)	N/R	N/R	N/R; all >65 years	N/R	N/R
Lester et al [38]	12 + 24; (B: Q2 2010; A: Q2s 2011-2013)	29,465	B: 6604; A: 22,861	<75 years; B: 5279 (80); A: 15,633 (68)	N/R	N/R
Ghibelli et al [45]	2 + 2; (B: 04 to 05/2012; A: 06 to 07/2012)	134	B: 74; A: 60	B: 81; A: 81	B: 27 (36); A: 25 (42)	B: 0 (0); A: 0 (0)
Stevens et al [39]	>6 + >12	N/R	N/R	N/R; all >65 years	N/R	N/R

[a]C: comparator group.

[b]I: intervention group.

[c]N/R: not reported.

[d]B: before.

[e]A: after.

[f]N/A: not applicable.

A total of 233,144 participants were included and assessed in RCTs (mean sample size: 21,199; range 196-72,072 participants). The crossover study included 3718 individuals. The pre-post intervention studies included more than 29,700 participants. However, some studies did not report a raw number of participants included in each study period. There was no information regarding whether missing data influenced the outcome assessment in eight studies (50%).

According to our inclusion criteria, all individuals were older than 65 years of age. The mean age in the selected studies was approximately 75 years. Females were often more prevalent, especially in larger studies.

The deprescribing target varied among the studies, and several papers used more than one criterion [30,32-34,40,45]. PIM was defined in some papers using internationally recognized criteria, such as the Beers Criteria (n=5) [32,34,39,40,45], the Screening Tool of Older People's Prescriptions (STOPP) criteria (n=3) [31,32,40], and the Anticholinergic Cognitive Burden Scale (n=1) [45]. In other studies (n=4), some group medications were specifically the target, such as benzodiazepines, opiates, and neuroleptics [35]; glyburide [36]; nonsteroidal anti-inflammatory drugs (NSAIDs), beta blockers, angiotensin-converting enzyme (ACE) inhibitors, or loop diuretics [41]; and diphenhydramine, metoclopramide, and antipsychotics [38].

Results of the Studies

The main results of the included studies are described in Tables 3 and 4. Several definitions and units were used to measure the impact of CDS tools on changes in PIP and PIM drugs (overall or concerning specific drugs). Studies assessed the following PIP- or PIM-related outcomes: number of PIMs started per 1000 visits [30], number of PIMs discontinued per 1000 visits [30], proportion of discontinued PIMs [30], percentage of PIMs [43], mean number of PIMs, risk of receiving a prescription for a drug exceeding the recommended maximum dose [42], risk of receiving a prescription for a drug exceeding the recommended standard doses [42], proportion of reconciliation errors corrected [32], proportion of recommendations implemented [32,33], proportion of patients with at least one PIM, and/or proportion of all prescribed medications that were PIM [33].

Table 3. Results of the included studies including changes in potentially inappropriate prescriptions or medications (N=16).

Study	PIP[a]- or PIM[b]-related outcomes	
	Changes in PIP or PIM drugs	Changes in specific PIP or PIM drugs
Randomized controlled trials		
Tamblyn et al [30]	Number of PIP started per 1000 visits C[c]: 52.2 vs I[d]: 43.8, RR[e] 0.82 (CI[f] 95% 0.69 −0.98); PIP discontinuation C: 44.5% vs I: 47.5%, RR: 1.14 (95% CI 0.98-1.33); number of PIP discontinued per 1000 visits C: 67.4 vs I: 71.4, RR 1.06 (95% CI 0.89-1.26)	Number of PIP started per 1000 visits: drug-disease contraindication C: 18.4 vs I: 16.6, RR 0.89 (CI 95% 0.72-1.10); drug-age contraindication C: 13.7 vs I: 10.7, RR 0.77 (CI 95% 0.59-1.00); excessive duration therapy C: 17.1 vs I: 13.3, RR 0.78 (CI 95% 0.61-0.99); therapeutic duplication C: 6.8 vs I: 6.1, RR 0.87 (CI 95% 0.69-1.11); number of PIP discontinued per 1000 visits: drug-disease contraindication C: 57.9 vs I: 62.6, RR 1.08 (CI 95% 0.85-1.36); drug-age contraindication C: 42.9 vs I: 40.7, RR 0.94 (CI 95% 0.79-1.13); excessive duration therapy C: 32.6 vs I: 32.3, RR 1.00 (CI 95% 0.77-1.29); therapeutic duplication C: 334.0 vs I: 317.1, RR 0.94 (CI 95% 0.59-1.51)
Price et al [31]	Change in PIP C: 0.1% vs I: 0.1%, *P*=.80	
Avery et al [41]	—[g]	At 6 months: history of peptic ulcer prescribed an NSAID[h] without a PPI/history of peptic ulcer without PPI[i] AOR[j] 0.58 (95% CI 0.38-0.89); asthma prescribed a β blocker/asthma AOR 0.73 (95% CI 0.58-0.91); aged ≥75 years long-term ACE[k] inhibitors or loop diuretics without urea and electrolyte monitoring in the previous 15 months aged ≥75 years receiving long-term ACE inhibitors or diuretics AOR 0.51 (95% CI 0.34-0.78); secondary outcomes AOR varied from 0.39-0.96; at 12 months: history of peptic ulcer prescribed an NSAID without a PPI/history of peptic ulcer without PPI AOR 0.91 (95% CI 0.59-1.39); asthma prescribed a β blocker/asthma AOR 0.78 (95% CI 0.63-0.97); aged ≥75 years receiving long-term ACE inhibitors or loop diuretics without urea and electrolyte monitoring in the previous 15 months aged ≥75 years receiving long-term ACE inhibitors or diuretics AOR 0.63 (95% CI 0.41-0.95); secondary outcomes AOR varied from 0.50-0.98
Erler et al [42]	CKD[l] patients with ≥1 prescription exceeding recommended maximum dose AOR 0.46 (95% CI 0.26-0.82); CKD patients with ≥1 prescription exceeding recommended standard dose by >30% AOR 0.66 (95% CI 0.36-1.21)	NS differences in the numbers of patients with potentially dangerous or contraindicated medications
Clyne et al [43]	Percentage of PIP I: 52% vs C: 77%, *P*=.02, AOR 0.32 (95% CI 0.15-0.70); mean number of PIP C: 1.18 vs I: 0.70, *P*=.02	Odds of PIP AOR 0.30 (95% CI 0.14-0.68); NS differences for duplicate or long-term benzodiazepines
Cossette et al [40]	Drug cessation or dosage decrease: at 48h C: 15.9% vs 45.8%, AD[m] 30.0% (95% CI 13.8-46.1); at discharge C: 27.3% vs I: 48.1%, AD 20.8% (95% CI 4.6-37.0); drug cessation: at 48h C: 15.1% vs 51.9%, AD 36.8% (95% CI 15.6-57.9); at discharge C: 34.4% vs I: 45.2%, AD 10.7% (95% CI −10.5 to 31.9); dosage decrease: at 48h C: 17.2% vs 38.1%, AD 20.9% (95% CI 4.1-45.8); at discharge C: 15.8% vs I: 52.4%, AD 36.6% (95% CI 12.3-60.9)	—
Fried et al [32]	Proportion of medication reconciliation errors corrected C: 14.3% vs I: 48.4%, *P*<.001; proportion of ≥1 TRIM recommendations implemented C: 21.9% vs I: 29.7%, *P*=.42	—
O'Sullivan et al [44]	Patients with ≥1 PIP C: 84.6% vs I: 82%	—
Terrel et al [33]	Proportion of visits with a PIP C: 3.9% vs I: 2.6, *P*=.02, OR[n] 0.55 (95% CI 0.34-0.89), ARR[o] 1.3% (95% CI 0.4-2.3); proportion of all prescribed medications that were PIP C: 5.4% vs I: 3.4, *P*=.006, OR 0.59 (CI 95% 0.41-0.85), ARR 2.0% (95% CI 0.7-3.3)	—

Study	PIP[a]- or PIM[b]-related outcomes	
	Changes in PIP or PIM drugs	Changes in specific PIP or PIM drugs
Raebel et al [34]	Newly dispensed ≥1 PIP rate per 100 patients C: 2.20 vs I:1.85, *P*=.002, RRR[p] 16%; newly dispensed ≥1 PIP only for indications included in intervention rate per 100 patients C:1.50 vs I: 1.10, *P*<.001	Newly dispensed ≥1 PIP rate per 100 patients: amitriptyline C: 0.61 vs I: 0.38, *P*<.001; chlordiazepoxide C: 0.05 vs I: 0.04, *P*=.55; diazepam C: 1.38 vs I: 1.28, *P*=.32; doxepin C: 0.14 vs I: 0.11, *P*=.24; flurazepam C: 0.01 vs I: 0.01, *P*=.69; ketorolac C: 0.00 vs I: 0.01, *P*=.50; meperidine (oral) C: 0.01 vs I: 0.01, *P*=N/A[q]; oxycodone/aspirin C: 0.00 vs I: 0.00, *P*=N/A; newly dispensed ≥1 PIP only for indications included in intervention, rate per 100 patients: amitriptyline C: 0.59 vs I: 0.37, *P*<.001; chlordiazepoxide C: 0.05 vs I: 0.04, *P*=.55; diazepam C: 0.71 vs I: 0.56, *P*=.002; doxepin C: 0.13 vs I: 0.09, *P*=.17; flurazepam C: 0.01 vs I: 0.01, *P*=.69; ketorolac C: 0.00 vs I: 0.01, *P*=.50; meperidine (oral) C: 0.01 vs I: 0.01, *P*=N/A; oxycodone/aspirin C: 0.00 vs I: 0.00, *P*=N/A; dispensings of chlorpropamide, hydrocodone/aspirin, or piroxicam C: 0 vs I: 0
Crossover studies		
Peterson et al [35]	Prescription recommended daily dose C: 19% vs I: 29%, *P*<.001; prescription orders with 10-fold dosing C: 5.0% vs I: 2.8%, *P*<.001; prescriptions in agreement with recommendation C: 18.6% vs I: 29.3%, *P*<.001; prescription of nonrecommended drugs C: 10.8% vs I: 7.6%, *P*<.001	Prescription orders with 10-fold dosing: benzodiazepines C: 3.5% vs I: 2.0%, *P*=.01; opiates C: 5.5% vs I: 2.8%, *P*<.001; neuroleptics C: 10.0% vs I: 7.5%, *P*=.35; prescriptions in agreement with recommendation: benzodiazepines C: 20.8% vs I: 28.2%, *P*<.001; opiates C: 16.6% vs I: 29%, *P*<.001; neuroleptics C: 22.5% vs I: 38%, *P*<.001
Pre-post intervention studies		
Ruhland et al [36]	—	Glyburide orders from total oral antidiabetic orders B[r]: 3.3% vs A[s]: 1.6%, *P*<.001; 17.8% patients transitioned off glyburide
Mattison et al [37]	Number of orders per total number of patients per day: not recommended medication B: 0.070 vs A: 0.054, *P*<.001; dose reduction medications B: 0.037 vs A: 0.037, *P*=.71; unflagged medications B: 0.033 vs A: 0.030, *P*=.03; number of orders per number of new patients per day: not recommended medication B: .333 vs A: 0.263, *P*<.001; dose reduction medications B: 0.182 vs A: 0.186, *P*=.51; unflagged medications B: 0.158 vs A: 0.148, *P*=.08	—
Lester et al [38]	—	>65 years prescription rates of: diphenhydramine B: 26.9% vs A: 20%, *P*<.001; metoclopramide B: 16.7% vs A: 12.5%, *P*<.001; antipsychotics B: 8.8% vs A: 9.2%, *P*=.80; ≥65 years: no significant changes for diphenhydramine, metoclopramide, or antipsychotics
Ghibelli et al [45]	Proportion of patients exposed to PIM at discharge B: 37.8% vs A: 11.6%; mean number of PIM per patient at discharge B: 0.4 vs A: 0.1	Proportion of patients exposed to PIM at discharge: high-dose short-acting benzodiazepines B: 21.6% vs A: 6.7%; ticlopidine B: 5.4% vs A: 0.0%; digoxin B: 5. 4% vs A: 1.7%; doxazosin B: 1.3% vs A: 1.7%; clonidine B: 1.3% vs A: 0.0%

Study	PIP[a]- or PIM[b]-related outcomes	
	Changes in PIP or PIM drugs	Changes in specific PIP or PIM drugs
Stevens et al [39]	Average percentage of PIMs per month: site 1 B: 11.9 vs A: 5.1, *P*<.001; site 2 B: 8.2 vs A: 4.5, *P*<.001; site 3 B: 8.9 vs A: 6.1, *P*=.007; site 4 B: 7.4 vs A: 5.7, *P*=.04	—

[a]PIP: potentially inappropriate prescription.
[b]PIM: potentially inappropriate medication.
[c]C: comparator group.
[d]I: intervention group.
[e]RR: relative rate.
[f]CI: confidence interval.
[g]No data.
[h]NSAID: nonsteroidal anti-inflammatory drug.
[i]PPI: proton-pump inhibitor.
[j]AOR: adjusted odds ratio.
[k]ACE: angiotensin-converting enzyme.
[l]CKD: chronic kidney disease.
[m]AD: absolute difference.
[n]OR: odds ratio.
[o]ARR: absolute risk reduction.
[p]RRR: relative risk reduction.
[q]N/A: not applicable.
[r]B: before.
[s]A: after.

Table 4. Results of the included studies including number of prescriptions, adverse drug reactions, and potential drug-drug interactions (N=16).

Study	Overall number of prescriptions	Adverse drug reaction	PDDI[a]	Others
Randomized controlled trials				
Tamblyn et al [30]	__ [b]	—	Number of PDDI started per 1000 visits C[c]: 1.5 vs I: 1.6, RR[d] 1.12 (CI[e] 95% 0.68-1.87); number of PPDI discontinued per 1000 visits C: 68.6 vs I[f]: 51.5 per 1000 visits, RR 1.33 (CI 95% 0.90-1.95)	Physicians with more computer problems downloaded information less often (*r*=−.31)
Price et al [31]	—	—	—	Description of 12 data quality probes; alert awareness: all participants in I were aware of STOPP[g] alerts, but not consistently; workflow and display: location on screen and workflow identified as barriers; study disruptiveness: considered as minimal
Avery et al [41]	—	—	—	Mean ICER[h] of intervention: at 6 months £65.6 (2.5-97.5 percentile 58.2-73.0); at 12 months £66.5 (2.5-97.5 percentile 66.8-81.5)
Erler et al [42]	—	—	—	—
Clyne et al [43]	—	—	—	Beliefs about Medicine Questionnaire AOR[i] 0.16 (CI 95% −1.85 to 1.07); 12-item Well-Being Questionnaire AOR −0.41 (95% CI −0.80 to 1.07)
Cossette et al [40]	—	—	—	LOS[j] (median, IQR[k]) C: 9.5 (5-21) vs I: 10 (6-19), *P*=.9; in-hospital death C:11 (8.6%) vs I: 6 (4.8%), *P*=.3; 30-day post discharge ER visits C: 27 (21.1%) vs I: 27 (21.4%); 30-day postdischarge readmissions C: 28 (21.9%) vs I: 20 (15.9%), *P*=.3
Fried et al [32]	Mean number of medications per patient C: 13.8 vs I: 13.3, *P*=.65	—	—	Mean patient active participation C: 2.7 vs I: 5.5, *P*=.001; percentage of patients assessment of care for chronic conditions score >10 C: 15.6% vs I: 29.7%, *P*=.06, OR[l] 2.73 (CI 95% 0.82-9.08); patient medication-related; communication C: 3.6 vs I: 7.5, *P*<.001; mean clinician facilitative communication C: 0.67 vs I: 1.53, *P*=.02; mean clinician medication-related communication C: 4.6 vs I:7.3, *P*=.002; percentage >1 recommendations C: 32.8% vs I: 63.6%, *P*<.001; OR 3.33 (95% CI 1.37-8.04)
O'Sullivan et al [44]	Total number of medications C: 3747 vs I: 4192, *P*<.001; median (IQR) number of medications per patient C: 9 (7-12) vs 12 (8-15), *P*<.001; number (%) of people with polypharmacy (≥5 medications); C: 346 (92.0) vs I: 346 (95.8), *P*=.44	Patients with ≥1 ADR[m] C: 20.7% vs I: 13.9%, *P*= 0.02, ARR[n] 6.8% (95% CI 1.5-12.3); RRR[o] 33.3% (95% CI; 7.7-51.7); NNT[p] 15 (95% CI 8-68)	—	CDS[q] alerts 1000 in 296/361 patients; intervention group attended 54.8% of recommendations; median (IQR) LOS days C: 9 (5-16) vs I: 8 (5-13.5), *P*=.44; hospital mortality C: 4.5% vs I: 4.7%, *P*>.05; interrater reliability for application of WHO-UMC[r] ADR causality criteria k= 0.81; Hallas ADR preventability criteria k= 0.87; application of Hartwig ADR; severity criteria k=0.56
Terrel et al [33]	—	—	—	CDS alerts 114 during 107 visits; 43% of recommendations accepted
Raebel et al [34]	—	—	—	—
Crossover studies		—	—	—

Study	Overall number of prescriptions	Adverse drug reaction	PDDI[a]	Others
Peterson et al [35]	Median (IQR) orders per admission C: 2 (1-3) vs I: 4 2 (1-3), *P*=.43	—	—	Number of altered mental status per 100 patient-days C: 21.9 vs I: 20.9, *P*=.17; median (IQR) LOS days C: 4 (2-6) vs I: 4 (2-6), *P*=.43; in-hospital fall rate C: 0.64 vs I: 0.28; falls per 100 patient-days, *P*<.001, AOR 0.50 (95% CI 0.30-0.82); fall injuries per 100 patient-days rate C: 0.17 vs I: 0.06, *P*=.09
Pre-post intervention studies		—	—	—
Ruhland et al [36]	—	—	—	CDS tool alerted 101 times for 75 providers during encounters for 76 patients over 90 days; physicians were more likely to transition patients off glyburide vs other health care providers (46.2% vs 8.0%, *P*<.001)
Mattison et al [37]	—	—	—	—
Lester et al [38]	—	—	—	—
Ghibelli et al [45]	—	—	Proportion of patients exposed to PDDI at discharge B[s]: 87.8% vs A[t]: 88.3%; mean number of PDDI per patient at discharge B: 4.5 vs A: 3.7	Median anticholinergic burden at discharge B: 1.5 vs A: 1.1
Stevens et al [39]	—	—	—	—

[a]PDDI: potential drug-drug interactions.
[b]No data.
[c]C: comparator group.
[d]RR: relative rate.
[e]CI: confidence interval.
[f]I: intervention group.
[g]STOPP: Screening Tool of Older People's Prescriptions.
[h]ICER: incremental cost-effectiveness ratio.
[i]AOR: adjusted odds ratio.
[j]LOS: length of stay.
[k]IQR: interquartile range.
[l]OR: odds ratio.
[m]ADR: adverse drug reaction.
[n]ARR: absolute risk reduction.
[o]RRR: relative risk reduction.
[p]NNT: number needed to treat.
[q]CDS: computerized decision support.
[r]UMC: Uppsala Monitoring Centre.
[s]B: before.
[t]A: after.

Effects of Interventions

The CDS tools consistently reduced the number of PIPs started and the mean number of PIPs per patient, while also increasing PIM discontinuation and drug appropriateness. However, in several cases statistical significance was not achieved for some of the assessed measures, such as for PIM discontinuation in the Tamblyn et al article [30], for change in PIMs in the Price et al study [31], and other studies described in Table 3.

Number of Prescriptions

With regard to the impact on the number of prescriptions, the RCT described by Fried et al [32] reported no significant reduction in the mean number of prescriptions in the group exposed to two Web apps. One study obtained information on medications and chronic conditions from an electronic health record, and the second study used an interface for data chart review, a telephone-based patient assessment, a set of automated algorithms evaluating medication appropriateness, and a

patient-specific medication management feedback report for the clinician. In a crossover study [35], there were no significant differences in the median number of medications prescribed per patient during the periods in which guided dosing of psychotropic medication was integrated into the Brigham Integrated Computer System.

In contrast, the RCT described by O'Sullivan et al [44] demonstrated that those in the intervention group (using CDS software structuring pharmacist review of medications designed to optimize geriatric pharmaceutical care) prescribed significantly fewer drugs (both total and median number of drugs). However, no impact was observed for the proportion of people with polypharmacy prescribed more than five drugs at once. This RCT was the only one addressing adverse drug reactions and it concluded that using this software significantly reduced the risk of adverse drug reactions. Furthermore, only 15 patients' medications needed to be reviewed to prevent one adverse drug reaction.

Number of Potential Drug-Drug Interaction

Only two studies assessed whether CDS tools could decrease the number of potential drug-drug interactions [30,44]. One CDS used in an RCT was found to decrease the initiation of PIP, but it did not have a similar impact on deprescription [30].

One pre-post intervention study observed that the proportion of patients exposed to potential drug-drug interactions increased after implementing a computer-based app that collects, stores, and automatically provides drug information to reduce or prevent PIPs [45]. However, the mean number of potential drug-drug interactions per patient at discharge was reduced. Statistical significance was not reported.

Other Measures

Other miscellaneous measures were reported in the studies examined, which should be highlighted. One RCT concluded that having computer problems was directly linked with PIP or PIM information download, and these computer problems could have an impact on the success of CDS tools [30]. Only one study described data quality probes; it found that professionals included in the intervention group were aware of STOPP alerts, although not in a consistent manner. Furthermore, the layout and impact on the workflow of the CDS tool were potential barriers to successful adherence [31].

Adherence to Computerized Decision Support Tools

Several RCTs reported the frequency of adherence to CDS recommendations by a health professional, with values ranging from 33% to 55% [32,33,44]. No significant reduction in the length of stay or intrahospital mortality was found in the RCT described by O'Sullivan et al [44]; in the Cosstte et al study [40], the differences between the intervention and control groups were not statistically different. Similarly, a crossover study found no difference in the length of stay between periods when the CDS tool was either active or inactive [35]. Likewise, no difference was observed with respect to patients' altered mental status or fall injuries. However, there was a significant decrease in the in-hospital rate.

The TRIM RCT concluded that the use of CDS tools significantly improved patients' active participation and facilitated communication between the clinician and the patient [32]. Another RCT found no significant impact on the Beliefs about Medicine Questionnaire or the 12-item Well-Being Questionnaire when general practitioners had access to information from a pharmacist and a medical review with Web-based pharmaceutical treatment algorithms and leaflets in addition to the usual care and simple, patient-level PIP postal feedback [43].

Cost-Effectiveness of Computerized Decision Support Tools

The cost-effectiveness of CDS tools was addressed in one RCT. The authors reported that there was a 95% probability that adding a pharmacist-led information technology complex intervention, in addition to computer-generated simple feedback, could be cost-effective, resulting in a willingness to pay £75 per error avoided at 6 months [41].

Risk of Bias in the Studies Examined

The RCTs received a total score according to the Cochrane Collaboration Risk of Bias tool that ranged from 1 [30,31] to 5 [41,43]. The procedure to guarantee allocation concealment was unclear in eight of ten RCTs. Complete blinding of participants and personnel was not possible due to the nature of the intervention. Blinding for the outcome assessment was not conducted in five studies [31,34,40,41,44], and was unclear if it was successful in another two [30,42]. Both of these biases may have resulted in an overestimate of the CDS tools' impact on PIP or PIM reduction (see Table 5).

Table 5. Risk of bias assessment (according to Cochrane Collaboration Risk of Bias tool) for the randomized controlled trials (n=10).

Study	Risk of bias items							Total score (max=7)
	Random sequence generation	Allocation concealment	Blinding of participants and personnel	Blinding of outcome assessment	Incomplete outcome data	Selective reporting	Other bias	
Tamblyn et al [30]	?[a]	?	–[b]	?	?	+[c]	–	1
Price et al [31]	+	?	–	–	?	?	–	1
Avery et al [41]	+	+	–	–	+	+	+	5
Erler et al [42]	+	?	–	?	+	+	–	3
Clyne et al [43]	+	?	–	+	+	+	+	5
Cossette et al [40]	+	?	–	–	–	-	+	2
Fried et al [32]	–	–	–	+	+	+	?	3
O'Sullivan et al [44]	?	?	–	–	+	+	–	2
Terrel et al [33]	+	?	–	+	?	+	–	3
Raebel et al [34]	+	?	–	–	?	+	+	3

[a]?: unclear risk of bias.

[b]–: high risk of bias.

[c]+: Low risk of bias.

Several studies did not report whether outcome data were available for all the participants included (n=4) [30,31,33,34]. Other biases were also found in five of the RCTs; namely, selection bias, performance bias, contamination, and underpowered sample sizes.

Regarding the pre-post intervention studies [36-39,45], they were considered high risk following the Cochrane Effective Practice and Organisation of Care [46]. For example, it is expected that pre-post intervention studies are more prone to the Hawthorne effect [47]. The Hawthorne effect happens when people (in this case, prescribers and patients) know they are being watched, which may lead to changes in behavior [47]. We consider that it is possible that being aware of one's study participation could have resulted in prescribers taking more care when prescribing medications.

Limited generalizability was also pointed out by several authors as a major limitation due to the context—single-center design—and the use of CDS tools that were created specifically for the study, which may not be available in other institutions.

Discussion

Principal Results

Despite the fact that withdrawal of PIPs is considered to be evidence-based [48], it is not an easy task [49]. CDS tools may play a role in supporting deprescription. From the 16 studies examined in this review, 10 were RCTs. Although RCTs represent stronger evidence, they lacked important data pertaining to clinical outcomes and presented a significant risk of bias (the total score of the studies using the Cochrane Collaboration Risk of Bias tool ranged from 1 to 5 with a mean value of 3). The most frequent biases included no blinding of health professionals and an unclear risk of breaking allocation concealment. If prescribers are not blinded, this can easily affect the deprescribing process. Health professionals may have been more susceptible to accepting the CDS tool recommendations. Alternatively, patients may have been more likely to agree with the withdrawal process. If a break in allocation concealment occurred, it is expected that investigators may have potentially included older adults that they considered best suited for the intervention group. Both types of bias may have led to an overestimation of the benefit of CDS tools.

We have also included five pre-post intervention studies. The nonrandomized nature of these studies is the major limitation of this analysis. The impact of CDS tools may be confounded by other changes that may have occurred in the institutions during the study periods.

We observed that almost two-thirds of the included studies were performed in the United States, and one-third were performed in European countries. This reflects the importance that has been given to this topic only in developed countries where electronic health record systems are widely available.

Overall Applicability and Quality of the Evidence

Seven studies were conducted in teaching hospitals and clinics [33,36-38,40,44,45], which may indicate potential bias. Teaching units are more prone to accept interventions in patient care, such as changes in a prescription through the use of CDS tools. We can assume that these professionals may be more likely to change a patient's prescription and, therefore, to address PIPs. This tendency may result in an overestimate of the impact of the intervention, and we can only speculate as to what would be the impact in a nonteaching unit.

There is a balance between the number of studies conducted in primary care versus secondary care institutions, and only one was conducted in both. The impact of CDS on PIP or PIM reduction was similar between settings despite differences in the health professional and population characteristics. This

suggests that the CDS tool might be successful in the context of a larger patient population.

The generalization of our results may be limited for several reasons. First, most studies used standard care as a comparator without providing additional details. In such a complex context, the management of older patients in institutions with several levels of care may mean that standard care could differ greatly between studies.

Second, the intervention varied greatly as a result of using different electronic systems, contents, and layouts. The intervention frequently included several features beyond the creation and application of a CDS tool itself.

Third, the main outcome definition was also diverse. Several studies used STOPP [31,32,40] and Beers Criteria [32,34,39,40,45] to define which medications were targeted. Both criteria are widely used worldwide, and although they do not provide a list of prohibited medications, they are an important tool for physicians due to their evidence-based rationale and constant updating. Nevertheless, the authors chose different groups of criteria for their outcome measures.

Fourth, the studies selected different participants and had widely variable sample sizes. Only two studies addressed potential drug-drug interactions [30,45] and one addressed adverse drug reactions [44]. Due to the increase of polypharmacy in older adults, the risk is higher for experiencing drug-drug interactions and adverse drug reactions. For the former, no significant impact was found, whereas for the latter, using a CDS tool significantly decreased the number of adverse drug reactions.

This tool, which included a clinical decision support software and a structured pharmacist review of medication [44], seems to be promising for aiding medication reconciliation activities. Most of the reconciliation issues highlighted by this CDS tool were accepted by the health care professionals involved. In particular, the Erler et al study [42] should, in our opinion, have assessed these two topics because they studied a population with renal impairment, which is particularly susceptible to adverse drug reactions and drug interactions. Similarly, only two studies assessed the impact of CDS tools on length of stay [35,40], and two assessed intrahospital mortality [40,44]. No differences were found between those using a CDS tool and those not using a CDS tool. Cost-effectiveness was also assessed by one study, which reported a 95% probability of a CDS tool being cost-effective due to a willingness to pay £75 to prevent an adverse drug reaction in a 6-month period [41]. The study's results may have been underestimated due to low adherence to CDS recommendations. Three RCTs that evaluated adherence reported values fluctuating from 33% to 55% [32,33,44]. Finally, we consider the possibility that the Avery et al trial [41] could have explored the issue of prescription NSAIDs to patients with a history of asthma as a secondary outcome because the authors had information on both conditions (prescriptions of NSAIDs and a history of asthma). This analysis could yield interesting information about the patterns of prescribing NSAIDs to these patients.

Strengths and Limitations

This review presents some limitations. We have chosen to include both RCTs (n=10) and pre-post studies (n=6). We acknowledge that the latter provide a lower level of evidence. Nevertheless, they have assessed some outcomes for which no additional evidence exists. In addition, we have focused our search on articles having PIP modification outcomes, thus some studies assessing changes in PIM may have been missed.

Our search terms were more limited to PIP; therefore, this paper may have missed some studies regarding PIM. Nevertheless, no new articles were found when searching in the references from the included studies and in the grey literature

Major strengths of our study include the fact that we have followed the Cochrane Collaboration Handbook [50], which makes our study less susceptible to major biases and errors. Furthermore, no new references were found from searches in the grey literature, pertinent scientific meeting books of abstracts, and the included studies' list of references, which suggests that our search strategy was exhaustive and all pertinent articles had been included.

However, the quality of the results of a systematic review is dependent on the available data. For all that was previously described, we believed that conducting a meta-analysis was not possible. Thus, only a narrative synthesis has been provided.

Comparison With Prior Work

To our knowledge, there are three previously published systematic reviews assessing the impact of CDS tools on PIP or PIM [51-27]. Due to an increase in the search period, the use of broader search criteria, and our overall methodology, we were able to include five additional RCTs [31,32,40,43,44]. These studies added evidence with new outcomes, such as well-being and patients' beliefs [43], reduction of adverse drug reactions [44], and users' perspectives [31].

The highlight of the findings in the more recent RCTs were as follows. In the study by Price et al [31], alerts with specific STOPP guideline content in electronic medical records positively changed PIPs (comparator: 0.1% versus intervention: 0.1%, P=.80), but not significantly. In the study by Clyne et al [43], the intervention consisted of Web-based pharmaceutical treatment algorithms that led to a lower percentage of PIPs (intervention: 52% versus comparator: 77%, P=.02). In the trial by Cossette et al [40], a computerized alert system-based pharmacist-physician intervention was able to significantly increase drug cessation or decrease dosage at discharge (comparator: 27.3% versus intervention: 48.1%; absolute difference 20.8%, 95% CI 4.6-37.0). In the TRIM trial [32], the proportion of medication reconciliation errors was significantly diminished (comparator: 14.3% versus intervention: 48.4%, P<.001). In the article by O'Sullivan et al [44], clinical decision support software reduced adverse drug reactions among older patients (control patients: 20.7% versus intervention patients: 13.9%, P=.02). In sum, articles published since 2012 substantiated the value of CDS to improve PIP- or PIM-related outcomes.

Conclusions

The use of CDS tools had a positive impact on PIP independently of the outcome definition in the majority of the studies included in our analysis. However, statistical significance was not always achieved. Several possible sources of bias and experimental limitations were found in the included studies, and evidence is lacking regarding the impact of CDS tools in potential drug-drug interactions, adverse drug reactions, length of stay, mortality, and cost-effectiveness.

This research suggests that RCTs assessing the impact of CDS tools could be conducted in both primary and secondary health care settings using medication targets defined by Beers or STOPP criteria.

To replicate the intervention in different RCTs, a standard CDS tool could be developed. These CDS tools could promote communication between physicians and pharmaceutical servives. These RCTs could also assess adverse drug reactions, quality of life measurements, and patient and professional satisfaction, with a reasonable follow-up to clarify the clinical usefulness of these tools.

Acknowledgments

This article was supported by National Funds through FCT-Fundação para a Ciência e a Tecnologia within CINTESIS, R&D Unit (reference UID/IC/4255/2019).

Conflicts of Interest

None declared.

References

1. OECD Data. 2019. Elderly population indicator URL: https://data.oecd.org/pop/elderly-population.htm#indicator-chart [accessed 2019-05-20]
2. World Health Organization. 2019. Ageing and health URL: https://www.who.int/news-room/fact-sheets/detail/ageing-and-health [accessed 2019-05-12]
3. World Health Organization. World Report On Ageing And Health. Geneva: World Health Organization; 2019.
4. World Health Organization. 2019. The global strategy and action plan on ageing and health URL: https://www.who.int/ageing/global-strategy/en/ [accessed 2019-06-22]
5. Prazeres F, Santiago L. Prevalence of multimorbidity in the adult population attending primary care in Portugal: a cross-sectional study. BMJ Open 2015 Sep 25;5(9):e009287 [FREE Full text] [doi: 10.1136/bmjopen-2015-009287] [Medline: 26408832]
6. Formiga F, Ferrer A, Sanz H, Marengoni A, Alburquerque J, Pujol R, Octabaix study members. Patterns of comorbidity and multimorbidity in the oldest old: the Octabaix study. Eur J Intern Med 2013 Jan;24(1):40-44. [doi: 10.1016/j.ejim.2012.11.003] [Medline: 23186603]
7. Lu W, Wen Y, Chen L, Hsiao F. Effect of polypharmacy, potentially inappropriate medications and anticholinergic burden on clinical outcomes: a retrospective cohort study. CMAJ 2015 Mar 03;187(4):E130-E137 [FREE Full text] [doi: 10.1503/cmaj.141219] [Medline: 25646290]
8. Gnjidic D, Hilmer SN, Blyth FM, Naganathan V, Waite L, Seibel MJ, et al. Polypharmacy cutoff and outcomes: five or more medicines were used to identify community-dwelling older men at risk of different adverse outcomes. J Clin Epidemiol 2012 Sep;65(9):989-995. [doi: 10.1016/j.jclinepi.2012.02.018] [Medline: 22742913]
9. Payne RA, Avery AJ, Duerden M, Saunders CL, Simpson CR, Abel GA. Prevalence of polypharmacy in a Scottish primary care population. Eur J Clin Pharmacol 2014 May;70(5):575-581. [doi: 10.1007/s00228-013-1639-9] [Medline: 24487416]
10. Motter FR, Fritzen JS, Hilmer SN, Paniz É, Paniz VM. Potentially inappropriate medication in the elderly: a systematic review of validated explicit criteria. Eur J Clin Pharmacol 2018 Mar 27;74(6):679-700. [doi: 10.1007/s00228-018-2446-0]
11. Marengoni A, Angleman S, Melis R, Mangialasche F, Karp A, Garmen A, et al. Aging with multimorbidity: a systematic review of the literature. Ageing Res Rev 2011 Sep;10(4):430-439. [doi: 10.1016/j.arr.2011.03.003] [Medline: 21402176]
12. Weng M, Tsai C, Sheu K, Lee Y, Lee H, Tzeng S, et al. The impact of number of drugs prescribed on the risk of potentially inappropriate medication among outpatient older adults with chronic diseases. QJM 2013 Nov;106(11):1009-1015. [doi: 10.1093/qjmed/hct141] [Medline: 23836694]
13. Tommelein E, Mehuys E, Petrovic M, Somers A, Colin P, Boussery K. Potentially inappropriate prescribing in community-dwelling older people across Europe: a systematic literature review. Eur J Clin Pharmacol 2015 Dec;71(12):1415-1427. [doi: 10.1007/s00228-015-1954-4] [Medline: 26407687]
14. Petrovic M, van der Cammen T, Onder G. Adverse drug reactions in older people: detection and prevention. Drugs Aging 2012 Jun 01;29(6):453-462. [doi: 10.2165/11631760-000000000-00000] [Medline: 22642780]
15. Cherubini A, Ruggiero C, Gasperini B, Dell'Aquila G, Cupido MG, Zampi E, et al. The prevention of adverse drug reactions in older subjects. Curr Drug Metab 2011 Sep;12(7):652-657. [Medline: 21495975]

16. Reeve E, Gnjidic D, Long J, Hilmer S. A systematic review of the emerging definition of 'deprescribing' with network analysis: implications for future research and clinical practice. Br J Clin Pharmacol 2015 Dec;80(6):1254-1268 [FREE Full text] [doi: 10.1111/bcp.12732] [Medline: 27006985]
17. Haverhals LM, Lee CA, Siek KA, Darr CA, Linnebur SA, Ruscin JM, et al. Older adults with multi-morbidity: medication management processes and design implications for personal health applications. J Med Internet Res 2011;13(2):e44 [FREE Full text] [doi: 10.2196/jmir.1813] [Medline: 21715286]
18. Mira JJ, Navarro I, Botella F, Borrás F, Nuño-Solinís R, Orozco D, et al. A Spanish pillbox app for elderly patients taking multiple medications: randomized controlled trial. J Med Internet Res 2014;16(4):e99 [FREE Full text] [doi: 10.2196/jmir.3269] [Medline: 24705022]
19. Clyne B, Fitzgerald C, Quinlan A, Hardy C, Galvin R, Fahey T, et al. Interventions to address potentially inappropriate prescribing in community-dwelling older adults: a systematic review of randomized controlled trials. J Am Geriatr Soc 2016 Jun;64(6):1210-1222. [doi: 10.1111/jgs.14133] [Medline: 27321600]
20. Cooper JA, Cadogan CA, Patterson SM, Kerse N, Bradley MC, Ryan C, et al. Interventions to improve the appropriate use of polypharmacy in older people: a Cochrane systematic review. BMJ Open 2015 Dec 09;5(12):e009235 [FREE Full text] [doi: 10.1136/bmjopen-2015-009235] [Medline: 26656020]
21. Loganathan M, Singh S, Franklin BD, Bottle A, Majeed A. Interventions to optimise prescribing in care homes: systematic review. Age Ageing 2011 Mar;40(2):150-162. [doi: 10.1093/ageing/afq161] [Medline: 21262782]
22. Rankin A, Cadogan CA, Patterson SM, Kerse N, Cardwell CR, Bradley MC, et al. Interventions to improve the appropriate use of polypharmacy for older people. Cochrane Database Syst Rev 2018 Dec 03;9:CD008165. [doi: 10.1002/14651858.CD008165.pub4] [Medline: 30175841]
23. Gnjidic D, Le Couteur DG, Kouladjian L, Hilmer SN. Deprescribing trials: methods to reduce polypharmacy and the impact on prescribing and clinical outcomes. Clin Geriatr Med 2012 May;28(2):237-253. [doi: 10.1016/j.cger.2012.01.006] [Medline: 22500541]
24. Bates DW, Kuperman GJ, Wang S, Gandhi T, Kittler A, Volk L, et al. Ten commandments for effective clinical decision support: making the practice of evidence-based medicine a reality. J Am Med Inform Assoc 2003;10(6):523-530 [FREE Full text] [doi: 10.1197/jamia.M1370] [Medline: 12925543]
25. Payne TH. Computer decision support systems. Chest 2000 Aug;118(2 Suppl):47S-52S. [doi: 10.1378/chest.118.2_suppl.47s] [Medline: 10939999]
26. Martins CM, da Costa Teixeira AS, de Azevedo LF, Sá LM, Santos PA, do Couto ML, et al. The effect of a test ordering software intervention on the prescription of unnecessary laboratory tests-a randomized controlled trial. BMC Med Inform Decis Mak 2017 Feb 20;17(1):20 [FREE Full text] [doi: 10.1186/s12911-017-0416-6] [Medline: 28219437]
27. Clyne B, Bradley MC, Hughes C, Fahey T, Lapane KL. Electronic prescribing and other forms of technology to reduce inappropriate medication use and polypharmacy in older people: a review of current evidence. Clin Geriatr Med 2012 May;28(2):301-322. [doi: 10.1016/j.cger.2012.01.009] [Medline: 22500545]
28. Monteiro L, Maricoto T, Solha IS, Monteiro-Soares M, Martins C. Computerised decision to reduce inappropriate medication in the elderly: a systematic review with meta-analysis protocol. BMJ Open 2018 Jan 30;8(1):e018988. [doi: 10.1136/bmjopen-2017-018988]
29. Higgins JP, Altman DG, Gotzsche PC, Juni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ 2011 Oct 18;343(oct18 2):d5928-d5928. [doi: 10.1136/bmj.d5928]
30. Tamblyn R, Huang A, Perreault R, Jacques A, Roy D, Hanley J, et al. The medical office of the 21st century (MOXXI): effectiveness of computerized decision-making support in reducing inappropriate prescribing in primary care. CMAJ 2003 Sep 16;169(6):549-556 [FREE Full text] [Medline: 12975221]
31. Price M, Davies I, Rusk R, Lesperance M, Weber J. Applying STOPP guidelines in primary care through electronic medical record decision support: randomized control trial highlighting the importance of data quality. JMIR Med Inform 2017 Jun 15;5(2):e15 [FREE Full text] [doi: 10.2196/medinform.6226] [Medline: 28619704]
32. Fried TR, Niehoff KM, Street RL, Charpentier PA, Rajeevan N, Miller PL, et al. Effect of the tool to reduce inappropriate medications on medication communication and deprescribing. J Am Geriatr Soc 2017 Oct;65(10):2265-2271 [FREE Full text] [doi: 10.1111/jgs.15042] [Medline: 28804870]
33. Terrell KM, Perkins AJ, Dexter PR, Hui SL, Callahan CM, Miller DK. Computerized decision support to reduce potentially inappropriate prescribing to older emergency department patients: a randomized, controlled trial. J Am Geriatr Soc 2009 Aug;57(8):1388-1394. [doi: 10.1111/j.1532-5415.2009.02352.x] [Medline: 19549022]
34. Raebel MA, Charles J, Dugan J, Carroll NM, Korner EJ, Brand DW, et al. Randomized trial to improve prescribing safety in ambulatory elderly patients. J Am Geriatr Soc 2007 Jul;55(7):977-985. [doi: 10.1111/j.1532-5415.2007.01202.x] [Medline: 17608868]
35. Peterson JF, Kuperman GJ, Shek C, Patel M, Avorn J, Bates DW. Guided prescription of psychotropic medications for geriatric inpatients. Arch Intern Med 2005 Apr 11;165(7):802-807. [doi: 10.1001/archinte.165.7.802] [Medline: 15824302]
36. Ruhland DJ, Bellone JM, Wilkes E. Implementation and assessment of an ambulatory prescribing guidance tool to improve patient safety in the geriatric population. Consult Pharm 2017 Mar 01;32(3):169-174. [doi: 10.4140/TCP.n.2017.169] [Medline: 28270272]

37. Mattison ML, Afonso KA, Ngo LH, Mukamal KJ. Preventing potentially inappropriate medication use in hospitalized older patients with a computerized provider order entry warning system. Arch Intern Med 2010 Aug 09;170(15):1331-1336 [FREE Full text] [doi: 10.1001/archinternmed.2010.244] [Medline: 20696957]
38. Lester PE, Rios-Rojas L, Islam S, Fazzari MJ, Gomolin IH. Impact of computerized physician order entry alerts on prescribing in older patients. Drugs Aging 2015 Mar;32(3):227-233. [doi: 10.1007/s40266-015-0244-2] [Medline: 25752906]
39. Stevens M, Hastings SN, Markland AD, Hwang U, Hung W, Vandenberg AE, et al. Enhancing Quality of Provider Practices for Older Adults in the Emergency Department (EQUiPPED). J Am Geriatr Soc 2017 Jul;65(7):1609-1614. [doi: 10.1111/jgs.14890] [Medline: 28388818]
40. Cossette B, Éthier J, Joly-Mischlich T, Bergeron J, Ricard G, Brazeau S, et al. Reduction in targeted potentially inappropriate medication use in elderly inpatients: a pragmatic randomized controlled trial. Eur J Clin Pharmacol 2017 Oct;73(10):1237-1245. [doi: 10.1007/s00228-017-2293-4] [Medline: 28717929]
41. Avery AJ, Rodgers S, Cantrill JA, Armstrong S, Cresswell K, Eden M, et al. A pharmacist-led information technology intervention for medication errors (PINCER): a multicentre, cluster randomised, controlled trial and cost-effectiveness analysis. Lancet 2012 Apr 07;379(9823):1310-1319 [FREE Full text] [doi: 10.1016/S0140-6736(11)61817-5] [Medline: 22357106]
42. Erler A, Beyer M, Petersen JJ, Saal K, Rath T, Rochon J, et al. How to improve drug dosing for patients with renal impairment in primary care--a cluster-randomized controlled trial. BMC Fam Pract 2012 Sep 06;13:91 [FREE Full text] [doi: 10.1186/1471-2296-13-91] [Medline: 22953792]
43. Clyne B, Smith SM, Hughes CM, Boland F, Bradley MC, Cooper JA, OPTI-SCRIPT study team. Effectiveness of a multifaceted intervention for potentially inappropriate prescribing in older patients in primary care: a cluster-randomized controlled trial (OPTI-SCRIPT study). Ann Fam Med 2015 Nov;13(6):545-553 [FREE Full text] [doi: 10.1370/afm.1838] [Medline: 26553894]
44. O'Sullivan D, O'Mahony D, O'Connor MN, Gallagher P, Gallagher J, Cullinan S, et al. Prevention of adverse drug reactions in hospitalised older patients using a software-supported structured pharmacist intervention: a cluster randomised controlled trial. Drugs Aging 2016 Jan;33(1):63-73. [doi: 10.1007/s40266-015-0329-y] [Medline: 26597401]
45. Ghibelli S, Marengoni A, Djade CD, Nobili A, Tettamanti M, Franchi C, et al. Prevention of inappropriate prescribing in hospitalized older patients using a computerized prescription support system (INTERcheck(®)). Drugs Aging 2013 Oct;30(10):821-828. [doi: 10.1007/s40266-013-0109-5] [Medline: 23943248]
46. Cochrane Effective Practice and Organization of Care. EPOC resources for review authors URL: https://epoc.cochrane.org/resources/epoc-resources-review-authors [accessed 2019-06-22]
47. Indrayan A. Medical Biostatistics. Third edition. New York: Chapman And Hall; 2013.
48. The 2019 American Geriatrics Society Beers Criteria® Update Expert Panel. American Geriatrics Society 2019 updated AGS Beers Criteria® for potentially inappropriate medication use in older adults. J Am Geriatr Soc 2019 Apr;67(4):674-694. [doi: 10.1111/jgs.15767] [Medline: 30693946]
49. Anderson K, Stowasser D, Freeman C, Scott I. Prescriber barriers and enablers to minimising potentially inappropriate medications in adults: a systematic review and thematic synthesis . BMJ Open 2014 Dec 08;4(12):e006544 [FREE Full text] [doi: 10.1136/bmjopen-2014-006544] [Medline: 25488097]
50. Cochrane Training. 2019. Cochrane handbook for systematic reviews of interventions URL: https://training.cochrane.org/handbook [accessed 2019-06-22]
51. Clyne B, Fitzgerald C, Quinlan A, Hardy C, Galvin R, Fahey T, et al. Interventions to address potentially inappropriate prescribing in community-dwelling older adults: a systematic review of randomized controlled trials. J Am Geriatr Soc 2016 Jun;64(6):1210-1222. [doi: 10.1111/jgs.14133] [Medline: 27321600]
52. Loganathan M, Singh S, Franklin BD, Bottle A, Majeed A. Interventions to optimise prescribing in care homes: systematic review. Age Ageing 2011 Mar;40(2):150-162. [doi: 10.1093/ageing/afq161] [Medline: 21262782]

Abbreviations

CDS: computerized decision support
NSAID: nonsteroidal anti-inflammatory drugs
PDDI: potential drug-drug interaction
PIM: potentially inappropriate medications
PIP: potentially inappropriate prescriptions
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 06.07.19; peer-reviewed by G Signorelli, B Alper, C Alvarez, C Kanu, M Basit; comments to author 01.08.19; revised version received 12.09.19; accepted 23.09.19; published 14.11.19.

Please cite as:
Monteiro L, Maricoto T, Solha I, Ribeiro-Vaz I, Martins C, Monteiro-Soares M
Reducing Potentially Inappropriate Prescriptions for Older Patients Using Computerized Decision Support Tools: Systematic Review
J Med Internet Res 2019;21(11):e15385
URL: https://www.jmir.org/2019/11/e15385
doi:10.2196/15385
PMID:31724956

Original Paper

Investigating Software Requirements for Systems Supporting Task-Shifted Interventions: Usability Study

Pepijn Van de Ven[1*], PhD; Ricardo Araya[2*], PhD; Maria Clara P de Paula Couto[3*], PhD; Maiara Garcia Henrique[4*], BSc; Damien Meere[5*], PhD; Ana Vilela Mendes[4*], PhD; Tim J Peters[6*], PhD; Antônio Seabra[4*], PhD; Renato M Franzin[4*], PhD; Paula Carvalho Pereda[4*], PhD; Marcia Scazufca[4*], PhD

[1]Health Research Institute, HIST Cluster, University of Limerick, Limerick, Ireland
[2]King's College, London, United Kingdom
[3]Friedrich-Schiller University, Jena, Germany
[4]Universidade de São Paulo, São Paulo, Brazil
[5]BT Ireland, Limerick, Ireland
[6]University of Bristol, Bristol, United Kingdom
[*]all authors contributed equally

Corresponding Author:
Pepijn Van de Ven, PhD
Health Research Institute, HIST Cluster
University of Limerick
Main Building D2043
Limerick
Ireland
Phone: 353 61 202925
Email: pepijn.vandeven@ul.ie

Abstract

Background: There is a considerable shortfall in specialized health care professionals worldwide to deliver health services, and this shortfall is especially pronounced in low-middle-income countries. This has led to the implementation of task-shifted interventions, in which specific tasks are moved away from highly qualified health workers to health workers with less training. The World Health Organization (WHO) has published recommendations for such interventions, but guidelines for software and systems supporting such interventions are not included.

Objective: The objective of this study was to formulate a number of software requirements for computer systems supporting task-shifted interventions. As the treatment of mental health problems is generally considered to be a task for highly trained health care professionals, it poses interesting case studies for task-shifted interventions. Therefore, we illustrated the use of the identified software requirements in a mobile system created for a task-shifted depression intervention to be provided to older adults in deprived areas of São Paulo, Brazil.

Methods: Using a set of recommendations based on the WHO's guidance documentation for task-shifted interventions, we identified 9 software requirements that aim to support health workers in management and supervision, training, good relationship with other health workers, and community embeddedness of the intervention. These 9 software requirements were used to implement a system for the provision of a psychosocial depression intervention with mobile Android interfaces to structure interventions and collect data, and Web interfaces for supervision and support of the health care workers delivering the intervention. The system was tested in a 2-arm pilot study with 33 patients and 11 health workers. In all, 8 of these 11 health workers participated in a usability study subsequent to the pilot.

Results: The qualitative and quantitative feedback obtained with the System Usability Scale suggest that the system was deemed to have a usability of between *OK* and *Good*. Nevertheless, some participants' responses indicated that they felt they needed technical assistance to use the system. This was reinforced by answers obtained with perceived usefulness and ease of use questionnaires, which indicated some users felt that they had issues around correct use of the system and perceived ability to become skillful at using the system.

Conclusions: Overall, these high-level requirements adequately captured the functionality required to enable the health workers to provide the intervention successfully. Nevertheless, the analysis of results indicated that some improvements were required

for the system to be useable in a task-shifted intervention. The most important of these were better access to a training environment, access for supervisors to metadata such as duration of sessions or exercises to identify issues, and a more robust and human-error–proof approach to the availability of patient data on the mobile devices used during the intervention.

(J Med Internet Res 2019;21(11):e11346) doi:10.2196/11346

KEYWORDS

task-shifting; community health workers; depression; medical informatics

Introduction

Background

Health care systems worldwide, but especially those in low-middle-income countries (LMICs), struggle with the high demand for the specialized resources traditionally used in the delivery of health care interventions. To address the lack of specialized resources, the most common strategy followed has been that of task-shifting. This strategy involves moving specific tasks, where appropriate, from highly qualified health workers to health workers with less training and fewer qualifications, so as to make more efficient use of the available human resources [1]. Task-shifting has a long history in many guises with structured implementations in, for example, China and Thailand from the 1950s and 1970s onward, respectively [2,3]. In Africa, task-shifting has been used for various diseases and notably in response to the HIV/AIDS pandemic [4]. Task-shifting is not limited to LMICs. For example, Maier and Aiken found that of 39 countries covering Europe, the United States, Canada, Australia, and New Zealand, 27 countries made use of task-shifting from physicians to nurses [5]. In such settings, task-shifting may improve access to services or reduce their cost. In LMICs on the other hand, the use of task-shifting may have a more pronounced effect as it may allow the delivery of health services where this was not possible before because of the lack of human resources. Rather than shifting tasks to nurses, in LMICs, most of the task-shifted programs have been delivered by community health workers (CHWs) who are generally people of all ages, often community members, with no professional education but who receive a few months of training.

A review of task-shifting studies in LMICs found that task-shifting was a promising approach to efficiency improvements and the increased provision of services at a given quality and cost [6]. Notwithstanding the merits of this strategy, there have been problems with its implementation [6,7]. Between 1980 and 1990, many established task-shifted programs were discontinued because of poor implementation, resourcing issues, or the absence of lasting health outcome improvements [8]. In recognition of these and other challenges, the World Health Organization (WHO) collated a set of recommendations for successful implementation of task-shifting interventions [1]. Campbell and Scott [8] divided these recommendations into 5 categories and added a sixth as follows:

1. Strong management and supportive supervision
2. Appropriate selection of CHWs
3. Suitable training
4. Adequate retention and incentive structures for CHWs
5. Good relationship with other health care workers
6. Community embeddedness of personnel and intervention

Information and communications technology (ICT), though acknowledged as a useful tool in some task-shifting interventions [9], is overlooked in these recommendations as an opportunity to overcome some of the fundamental issues in supporting task-shifting. ICT can play an important role in the training, support, and supervision of CHWs delivering task-shifted interventions. The need for such functionality is illustrated by several studies of task-shifted interventions in mental health, which reported challenges around training and supervision of the intervention providers [10,11], treatment quality [12], and fidelity [13].

Objectives

This study sought to address this gap by formulating a set of software requirements to extend the requirements listed by Campbell and Scott [8] and illustrating the use of these requirements in a software platform for the support of a task-shifted depression intervention. Depression is common among older adults [14-17], impacts negatively on quality of life [18,19], has negative social and health consequences [20-22], and increases health care utilization and costs [23,24]. There are effective treatments for depression in later life [25-27], but these are complex interventions, which often require specialized resources. One of the most salient barriers to deliver these programs is the lack of specialized staff [28-32]. Hence, a task-shifted depression intervention is a challenging, potentially impactful demonstrator for the task-shifting software requirements we identified in this study. The intervention aimed to allow CHWs, who normally provide basic services to elderly citizens within Brazil's family unit-based health system [33], to deliver a complex psychosocial intervention to older adults residing in poor neighborhoods of São Paulo, Brazil. As far as we are aware, specific software requirements for ICT systems that play supporting roles in task-shifting interventions are not currently available. Although the app described focusses on mental health interventions, the applicability of the chosen approach is not limited to the domain of mental health.

Methods

Definition of Task-Shifting Software System Requirements

Four of the recommendation categories proposed by Campbell and Scott are relevant to the functionality that can be provided by ICT systems: management and supervision, training, good relationship with other health care workers, and community embeddedness. During the initial stages of the project, a needs analysis and *use case* scenarios were employed to define software requirements for these 4 categories. In this phase, a

psychiatrist and 3 psychologists were involved in the drafting of a number of use cases to describe a typical use of the system and the functionality required to support the CHWs in delivering the intervention. Subsequently, a nonfunctional prototype based on mock-ups was tested and discussed further, resulting in updated use cases and the first version of the requirement specification. The first version of the requirement specification formed the basis of a functional prototypes system that was presented to CHWs. Their feedback was incorporated in the final pilot system.

Software Requirements for Strong Management and Supportive Supervision

The mobile nature of modern ICT systems allows users to manage their tasks effectively and obtain appropriate supervision input as and when required. Such interactions can be user initiated, prompted by the system, or may result from supervisor intervention. The developed system contributes to such management and supervision by implementing the requirements described further.

Requirement 1: Guidance

The ICT system should provide the CHW with a means of structuring the intervention, automatically presenting the key aspects of the intervention at the appropriate time.

Requirement 2: Decision Support

The system should either aid in decision making, or automatically make certain decisions for the care provider.

Requirement 3: Supervision

The system should enable supervisors of the CHWs to remotely monitor the intervention, assess progress and potential issues, and initiate corrective action when required.

Requirement 4: Accountability

The system should provide all users with the trust and confidence that (1) they receive automated prompts in regard to their duties; (2) the recording of the execution of such duties is performed automatically insofar as practically and ethically possible; (3) any deviation from these duties is flagged accordingly and reported to line managers; and (4) any storage and communication of patient data are performed in a secure manner.

Requirement 5: Record Keeping

The system should automatically maintain records of information resulting from the intervention. These records should then be available for later sessions to support Guidance and Decision Support and for Supervision and Accountability purposes.

Software Requirements for Suitable Training

Characteristics of learning identified as important for task-shifting [34], such as being learner-centered, experiential, problem-orientated, and context-appropriate can be appropriately provided by ICT, especially as a refresher after initial, more formal training.

Requirement 6: Training Environment

The system should allow CHWs to quickly review the content of a specific session and practice delivery of the intervention and all its exercises.

Software Requirements for Good Relationship With Other Health Care Workers

The relationship between CHWs within a team and between CHWs and their supervisors is crucial to a successful task-shifted intervention [35]. ICT can contribute to a good relationship by providing effective and transparent communication between CHWs among themselves, and between CHWs and their supervisors or general practitioners.

Requirement 7: Automated Communication

The system should automatically communicate important information and events to relevant stakeholders. It should be clear to all stakeholders what information is relayed to whom. Where required, stakeholders should be able to respond.

Requirement 8: User-Initiated Communication

The system should allow immediate communication with relevant stakeholders involved in the care of that particular patient.

Software Requirements for Community Embeddedness of Personnel and Intervention

Campbell and Scott [8] defined community embeddedness as "...when community members 'own' the project by having substantial control over the selection, monitoring, activities and priority-setting of CHWs." Whereas this definition goes far beyond the physical location of the intervention, an important aspect of community embeddedness is that interventions can be provided where this is deemed most suitable by the community, and often this may be the patient's own home. Especially for the older adults targeted in this work, mobility may be a limiting factor in the delivery of care through health centers.

Requirement 9: Delivery in the Community

It should be possible to deliver the full intervention in the community, certainly in a location acceptable to the patient and if necessary in their own home. Hence, all intervention resources and patient data should be accessible on a mobile device without direct connection to the internet, for instance through an app.

Development of the Pilot System

The system requirements gathered in the early stages of this development process indicated the necessity for an architecture with interfaces for CHWs on mobile devices, a central storage of gathered data, and Web interfaces for supervisory actors. The resulting system, which was named PROACTIVE, contains a mobile Android app developed for tablets with a minimum screen size of 8″ and various interfaces accessible through a Web-based portal for remote supervision and management purposes.

The Psychosocial Intervention

One of the main goals of the intervention is to strengthen the autonomy of the patients and highlight the role they can play

in their own therapy. The intervention combines psychosocial techniques tailored to individual participants with embedded support mechanisms for nonspecialist health workers delivering the intervention. Behavioral activation was used as the main psychosocial approach in view of its demonstrated feasibility and efficacy for the treatment of depression [36]. It is a simple technique to apply, requires only a short period of professional training [37], and is suitable for delivery by nonspecialists [38,39]. The intervention further incorporates elements of psychoeducation (education about depression and simple coping strategies) and relapse prevention.

Mobile Android Interfaces

The CHW performs all interactions and sessions with the patient using the Android app. Upon logging in with a username password combination, the Android app's starting screen gives the user an overview of the patients in their care (see Figure 1).

When selecting a user and choosing *Start Session*, the correct session content will be compiled for the chosen user and an intervention session started. The various screens in each session can be navigated using swipe actions to move to neighboring screens similar to browsing through a book page by page. In addition, *chapters* in the session can be selected using a horizontal menu bar at the top of the screen, and pages within these chapters can be selected using a vertical menu at the left-hand side of the screen. Wherever the interface presents or requests dynamic patient or user-specific data, a custom-built application programming interface (API) is used to provide a quick and easy means of defining user interfaces with automated storage and retrieval of information. The information thus obtained is stored in database tables.

The first screen of every intervention session (apart from the first session) displays the homework assignments agreed upon in the previous session (Figure 2) and allows the patient and CHW to start the new session discussing and recording progress and exploring potential barriers and solutions for homework assignments.

After reviewing their homework, the patients will respond to a number of questionnaires. On every visit, depression is assessed with the 9-item Patient Health Questionnaire (PHQ-9) [40], and mood is rated on a Likert rating scale (Figure 3). To assist CHWs in this task, a video presented during the first session explains to patients why and how the PHQ-9 is conducted, and the CHWs and patients have access to these videos throughout all future sessions. On the basis of the answers provided in the PHQ-9, the system will, if any of the answers indicate the presence of a depressive symptom, automatically add an extra question to assess how PHQ-9 symptoms have affected the patient's life. Any positive response to the ninth question of the PHQ-9 on suicidal ideation automatically determines the immediate assessment of suicidal risk. For this, questions frequently used to assess immediate suicide risk (if there was any suicide attempt during the last 14 days and when was this attempt, and if the person has plans and means for another attempt) are shown on the screen. If the answer to any of these questions is positive, the CHW will immediately see instructions in the app on how to proceed in this situation, that is, to stay with the patient until a friend or family member has arrived. Other information, such as medication use (Figure 4), can be updated if required using a questionnaire.

Figure 1. Patient overview screen.

Figure 2. Homework review screen.

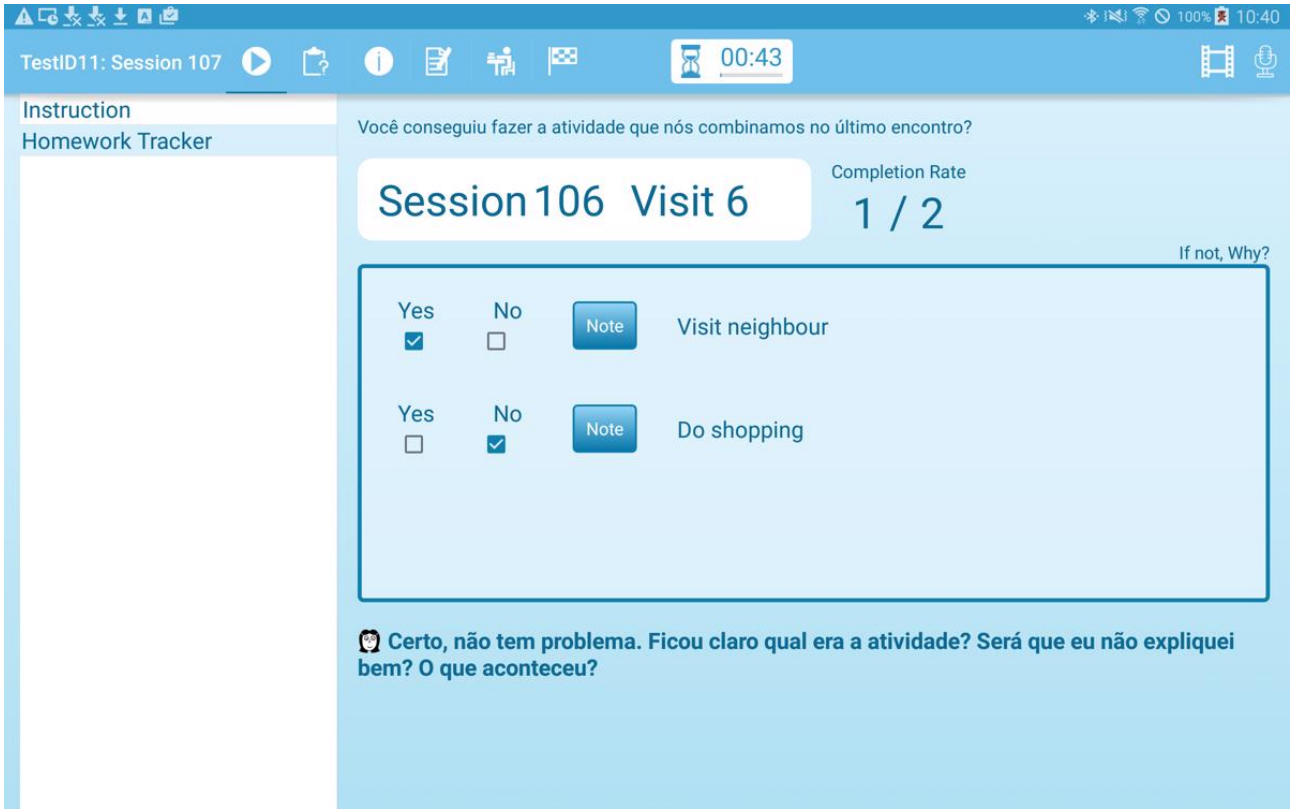
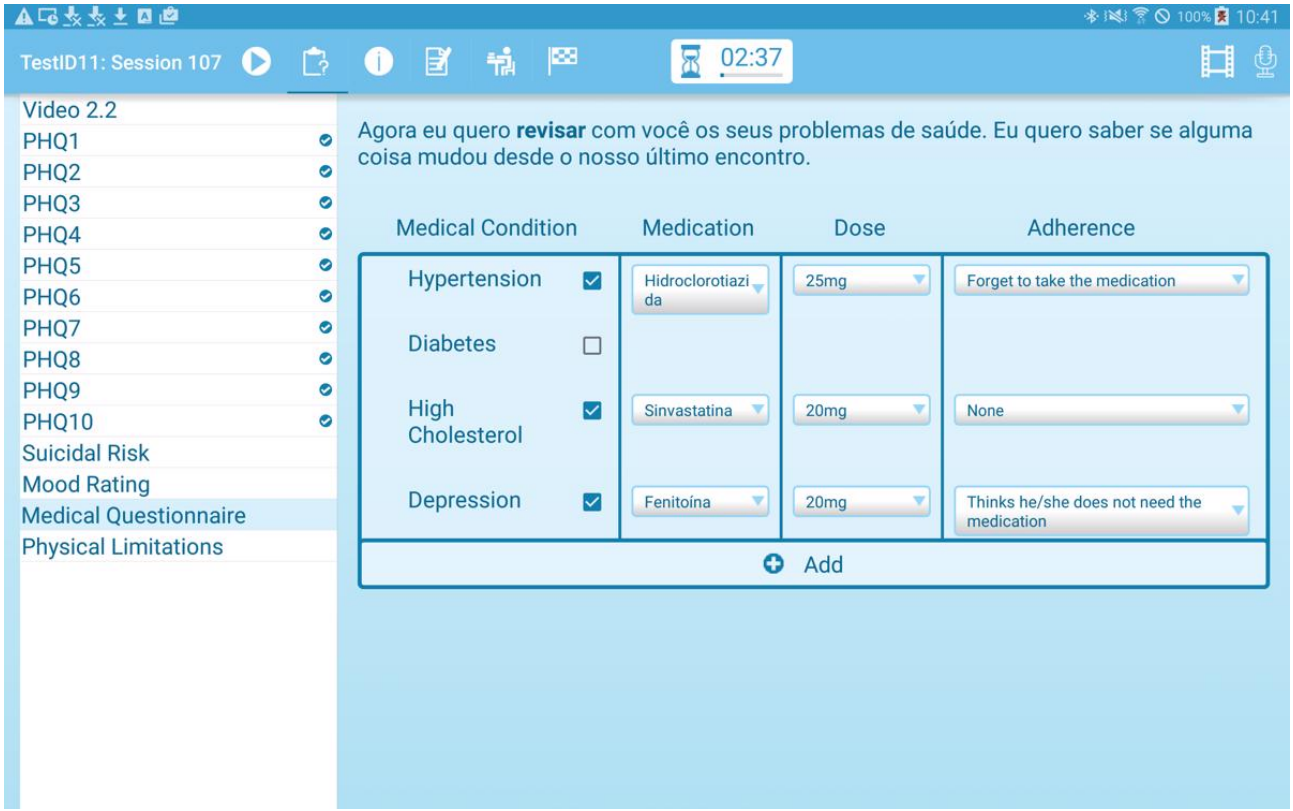


Figure 3. Mood rating screen.

Figure 4. Medication questionnaire.



Having gathered and updated the patient's mental and physical health status, the intervention continues with multimedia resources that explain aspects of the patient's symptoms, behaviors, and ways to improve these symptoms. In a series of videos specifically designed for this intervention, an actor explains the symptoms of depression, the vicious circle leading to a worsening of the depression and ways to counteract this vicious circle, the importance of doing pleasant activities to improve mood, how avoidant behaviors are linked to depression, how to plan pleasant (healthy) activities, and how to prevent relapse. As CHWs have limited training in the symptoms of depression and the principles underlying the intervention, these videos support the CHWs in providing patients with the required information on their depressive symptoms and the purpose of the intervention. One of the main goals of the intervention is to strengthen the autonomy of the patients and highlight the role they have in their own improvement. For this reason, in all sessions the patients do activities on the app, supported by the CHW, and plan activities to do between sessions. This strategy helps patients to develop the autonomy to identify and deal with symptoms of depression at present and in the future.

Sessions conclude with the scheduling of a new appointment. This scheduler recommends the date for the next appointment based on the patient's current progress in the therapy and structures the process of making an appointment for a new visit to complete the current session or to move on to the next session in the therapy. As internet connectivity is not always available at the patients' homes, the app contains all information for the patients under the care of the CHW who uses the device.

As mentioned previously, an important goal of the system is to empower CHWs with limited training in mental health service provision to provide a psychosocial intervention. To this end, the system provides automated support to the CHWs. This support consists of session content recommendations, as well as support in scheduling intervention visits and further recommendations personalized to that patient, such as warnings regarding adherence to the intervention, suicidal ideation, and progress in the intervention. The CHW has access to these recommendations and warnings in a dedicated screen in the app. Where necessary, the system forwards recommendations and warnings to supervisors by email.

To further support the supervision of CHWs, he or she can, in consultation with the patient, choose to make audio recordings of (parts of) the session for use in future intervention or supervision sessions. These audio recordings are stored on the device and are only accessible to the CHW.

Web Interfaces for Monitoring and Supervision

The Web interfaces allow individuals with supervisory roles access to the system. The information shown to these individuals varies slightly based on the objective to show only the most pertinent information. Those that are responsible for the running of the intervention at a high level (Trial Managers or their equivalent in the health system) see aggregate data on a per-patient basis, whereas clinical supervisors see more detailed information for each patient related to the clinical care in which one of their CHWs is involved.

Experimental Validation and User Experience Testing

The PROACTIVE system was tested in a pragmatic 2-arm pilot study in 2 family health units in the Northern area of São Paulo city, Brazil [41]. A previous survey of older inhabitants of São Paulo's poor neighborhoods found that only 12.3% of those

identified as having depression were receiving treatment [42]. The insufficient availability of mental health services to such populations has previously raised the alarm [43], and PROACTIVE is a first attempt to provide a simple, feasible, and affordable depression intervention to older inhabitants of São Paulo's poor neighborhoods. Eligible participants were aged 60 years and older. Exclusion criteria were not having depression as assessed by the PHQ-9 (PHQ-9<10); complete deafness; terminal illness; high risk of suicide; or incapability to communicate (for example, cognitive impairment and mental illness either relayed by a family member or detected by the researcher).

Before the start of the intervention, 8 CHWs and 3 nurse assistants (NAs) were trained and provided with an overview of the intervention, session contents, how to use the technological support platform, psychosocial techniques to deliver the intervention, and ways to engage with patients.

The training program consisted of 3 full days of intensive training delivered by 2 psychologists. During the intervention, the system was used for a total of 17 weeks by the 8 CHWs and 3 NAs who cared for a total of 33 patients with 15 individuals in a low-intensity (8 sessions) and 18 in a high-intensity group (11 sessions). Overall, 19 patients completed all sessions within the timeframe available. Completion rate in the low-intensity group was 13 out of 15 with completion in the high-intensity group 4 out of 18. The main reason for the low completion rate in the high-intensity group was the slightly curtailed time available for the follow-up in the pilot study. The mean PHQ-9 at follow-up (approximately 24 weeks after assignment to a CHW) was 12.3 (SD 3.7) and 3.8 (SD 3.9) in the control and intervention arms, respectively, and the follow-up rate was 92% (23/25) and 94% (31/33) in the control and intervention arms, respectively.

Upon completion of the pilot, the 8 CHWs and 2 of the 3 NAs involved in the intervention participated in a qualitative assessment of the system in which the System Usability Scale (SUS) [44] and the Technology Acceptance Model (TAM) [45] were used to elicit the CHW's perception of usability of the system. The NAs were excluded from this analysis, as their higher levels of training would result in optimistically skewed conclusions. The average age of the CHWs was 37.3 years (SD 6.7 years). Out of the 8 CHWs, 7 indicated they had smartphones and of those with smartphones, all used WhatsApp, 5 used Facebook, and 3 used their smartphones to play games. Half of the respondents had used tablets before their use of tablets in this study. They rated their level of difficulty with the use of technology on average as 1.9 on the scale provided in Multimedia Appendix 1, indicating that, on average, the respondents perceived their use of technology between *A lot of difficulty* and *Cannot handle technology*.

The questions and Portuguese translations of the SUS are shown in Multimedia Appendix 1, with responses provided on a 5-point Likert scale ranging from 0 to 4 with only the end points labelled as strongly disagree (coded as 0) and strongly agree (coded as 4).

The questions and Portuguese translations of the TAM questionnaire can be found in Multimedia Appendix 1. The responses to the TAM questionnaire are provided on a 7-point Likert scale ranging from extremely unlikely (coded as 1) to extremely likely (coded as 7) (see Multimedia Appendix 1).

For the analysis of results, we calculated the SUS score in the usual way, which involves summing the scores on all positive elements of the questionnaire (ie, all odd-numbered statements), summing the scores on all negative elements (even-numbered statements), and subtracting the latter sum from the first. By adding 20 to the result and subsequently multiplying the total by 2.5, the overall score will be on a scale of 0 to 100. Rather than interpreting the score as a percentage score, it should be seen as a percentile rank with an SUS score of 68 as the mean score [46].

Results from the TAM questionnaire were analyzed as the *perceived usefulness (PU)*, calculated as the mean score of questions 1 to 6 of the TAM questionnaire, and *ease of use (EoU)*, calculated as the mean score of the questions 7 to 12 of the TAM questionnaire.

Written informed consent was received from all participants in the study. The Ethical Committees of the Faculty of Medicine of the University of São Paulo (number 1.339.865) and of the Municipal Health Secretariat of São Paulo (number 1.340.790) approved this study. The trial was registered with the Registro Brasileiro de Ensaios Clínico (ReBEC), number RBR-5nf6wd-ReBEC.

Results

The mean overall SUS score was 65.6, which in previous studies has been associated with an overall rating of between *OK* and *Good* [47]. Mean ratings of the responses to the individual items are shown in Table 1. These show that, whereas agreement with the positive aspects of the SUS (the odd-numbered questions) is always higher than neutral (a score of 2), agreement with the negative aspects of the SUS questionnaire (the even-numbered items) is higher than neutral for questions 4 and 6. From the questionnaire wording, this indicated that users experienced aspects of the system as inconsistent and expected to need assistance for further use of the system.

The mean scores for *PU* and *EoU* were 5.4 (SD 1.7) and 5.6 (SD 1.3), respectively, which indicates mildly positive views between *somewhat likely* and *likely* on *PU* and *EoU*. Per-question average responses on the *PU* and *EoU* are shown in Table 2 and provide a better insight in aspects of the system that require improvements. The 2 aspects of PU that were scored lowest (PU4 and PU5) indicated that the respondents perceived the ability of PROACTIVE to help them be more effective at doing their job and making the job easier *somewhat likely*. Of note for *PU* is that the participants performed tasks with the app they had never performed before, and their frame of reference did not allow comparison with a past experience without the use of the app. In that respect, the *EoU* responses are of greater interest as they give an insight in usability experiences.

Table 1. Mean scores on individual items of the System Usability Scale.

System Usability Scale question	Mean score on question
1	3.1
2	1.4
3	2.6
4	2.5
5	3.5
6	2.1
7	2.6
8	1
9	3.1
10	1.8

Table 2. Average scores on the Technology Acceptance Model.

Technology Acceptance Model question	Mean score on question
1	6
2	5.6
3	5.4
4	5
5	5.1
6	5.9
7	5.4
8	5.1
9	6
10	5.9
11	5.1
12	6

In all, 3 items on this scale received mean scores below the overall mean EoU score of 5.6. These items relate to learning to operate (EoU1), using correctly (EoU2), and becoming skillful (EoU5) at use of the app.

To put this in perspective, Figures 5-7 show plots of the relationships between the user-reported difficulty in using technology versus SUS, PU, and *EoU*, respectively. These figures suggest that lower scores for SUS, PU, and EoU can partly be explained by the participants' level of difficulty with use of technology. This in turn suggests that the *PU* and *EoU* improvements may be realized through a combination of modifications to the app and the ability to gain more experience in its use. It is acknowledged that these results were obtained in a pilot study from a group of only 8 participants, thus limiting their statistical value.

Figure 5. System Usability Scale (SUS) versus technology experience.

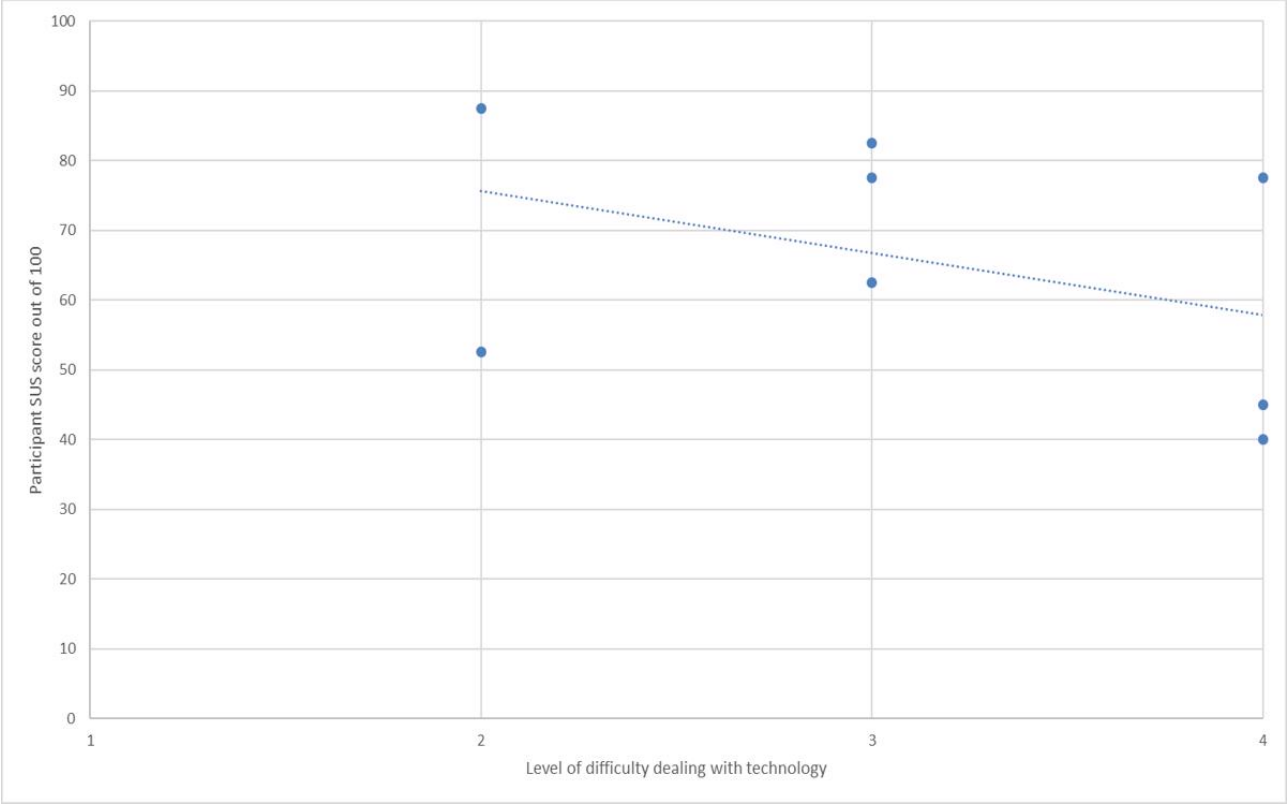


Figure 6. Perceived usefulness (PU) versus technology experience.

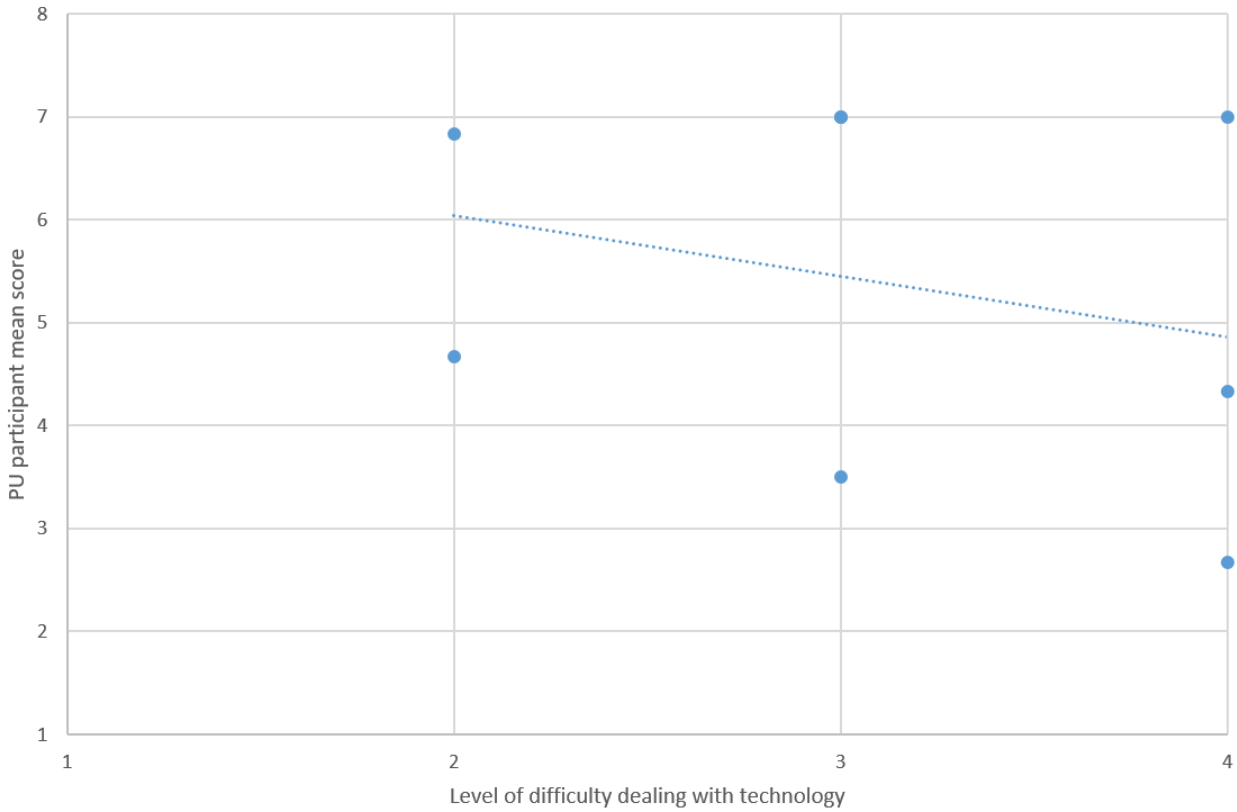
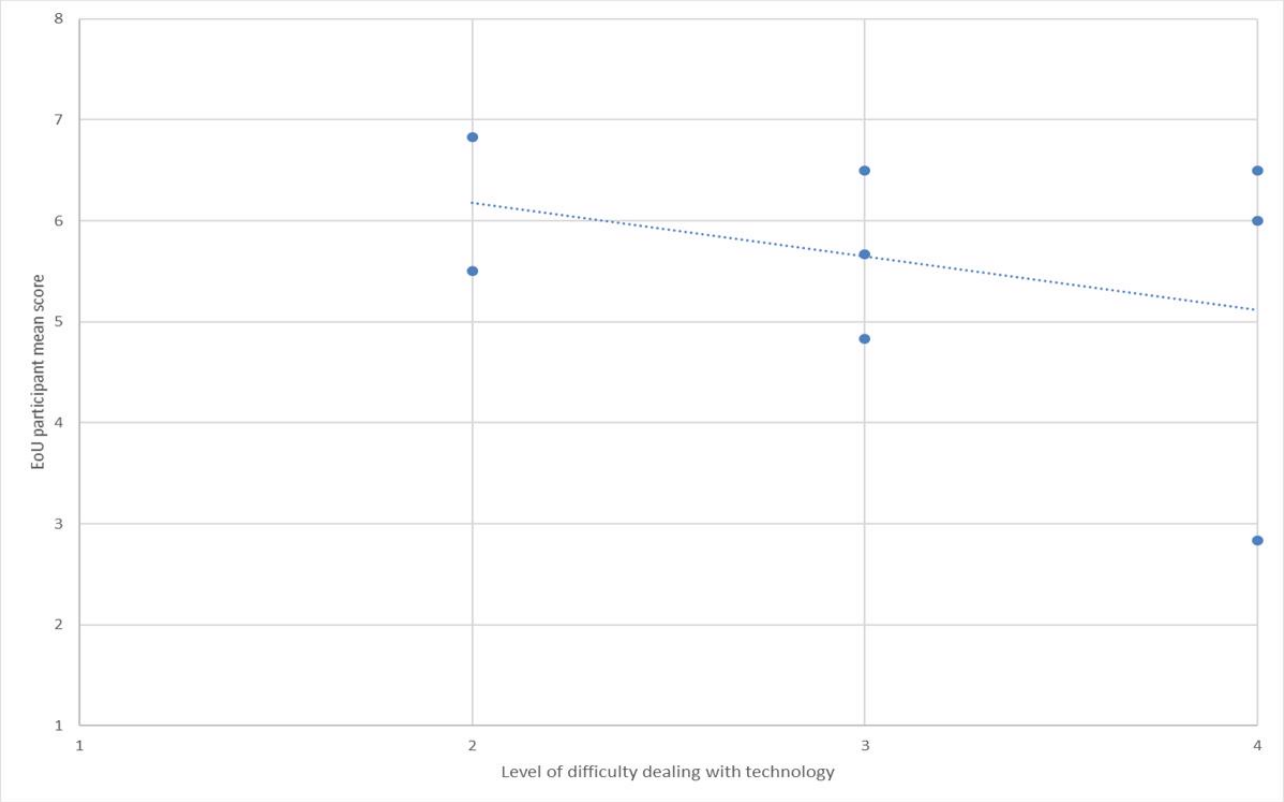


Figure 7. Ease of use (EoU) versus technology experience.



Discussion

Principal Findings

Feedback obtained during the pilot and the qualitative software validation performed with end users provided important insights in the extent to which the high-level requirements identified for the system were implemented successfully.

Requirement 1: Guidance

In the usability assessment, some of the users indicated that they found the system inconsistent and expected to require technical assistance during normal use of the system. Whereas these results may partly stem from bugs encountered and resolved during the pilot, they do indicate that further guidance is required. Solutions may be to hide part of the interface, such as the list of pages in a chapter displayed on the left-hand side of the screen by default and more extensive logging of use of the app, such that issues can be identified automatically. For example, longer than expected or usual activity on a particular screen of the app may indicate that users encountered issues with the content presented on the screen and such findings can be discussed in supervision meetings.

Requirement 2: Decision Support

Decision support was visible to end users in the form of recommendations on the course of action after 3 initial sessions resulting in the patient being allocated to a high-intensity (8 more sessions) or low-intensity (5 more sessions) intervention, questionnaires on suicidality, and advice to discuss patients that did not show progress (based on trends in PHQ-9 responses) with supervisors. The algorithm for allocation to the low-intensity or high-intensity treatment was not disclosed to the health workers, so they could not influence the allocation of patients to the second phase of treatment.

Requirement 3: Supervision

In a full-scale trial or clinical roll-out of this intervention, CHWs will be supervised by the clinical supervisors who normally have supervisory responsibilities for these CHWs. During the pilot, supervision was performed by research psychologists who conducted the pilot study. Data captured by the system were used to discuss the progress of patients and to make suggestions for future sessions.

The functionality to make audio recordings of sessions was added to the app during the pilot upon request. For simplicity and ethical considerations, it was decided not to send the resulting audio files to the server. Hence, reviews can only be done face-to-face with the CHW, who is the only person with access to the files created on the device. It would be worth considering whether the benefits that could be derived from making the audio files available on the server would weigh up against potential issues around privacy. It may also be possible to implement a consent mechanism for audio uploads to the server on a per-user or per-instance basis. A further consideration is that the storage of audio data would require more server storage capacity, but this should not be a limiting factor with modern servers.

Requirement 4: Accountability

Accountability functionality was enabled by the patient data gathered during intervention sessions, but also by logs relating to session timing and appointments. The app logged when a session was started and finished and kept track of appointments made with the patient. It also allowed the CHW to make note

of missed appointments and the reasons. These data proved invaluable during the pilot in several ways. For example, the patient responses gathered during sessions allowed automated content selection during interventions and the generation of patient-specific prompts to CHWs, thus providing robust support to the latter in following protocols and procedures. The data on scheduled, completed, and missed sessions provided valuable information on patient progress and potential barriers to their involvement in the intervention. For (trial) supervisors, the logged data were presented in such a way that these individuals obtained a high-level overview of the intervention delivered by CHWs and progress of their patients, thus allowing timely corrective action when required.

During the pilot, the geolocation of a session was not recorded and CHWs were also not prevented from accessing a session for a patient when there was no future appointment for that patient. Although this may be seen as overly rigid, there may be merit in locking the app for a user until shortly before the next scheduled appointment. This would prevent inadvertent data logging for a user (for example, because of briefly accessing the next session in the app when merely intending to review the session the next patient will receive) and would also encourage CHWs to maintain complete diaries of patient visits.

Treatment recommendations and warnings regarding a lack of progression or suicidality were displayed both in the app and in the supervisors' Web interface. For full accountability, a protocol should be implemented that requires CHWs and their supervisors to indicate that the recommendations and warnings have been reviewed and followed up.

Requirement 5: Record Keeping

The patient data logged in the sessions were used to personalize subsequent sessions, thus facilitating CHWs to provide a contextualized intervention. This, in turn, enabled sessions to be conducted efficiently (for example, the medical questionnaire can be reviewed and updated very quickly) and allowed an objective review of progress (through reviewing homework agreed on in a previous session and by allowing read-only access to tabular and graphical representations of PHQ-9 response, mood ratings, and homework completion on a session basis). In addition, aggregate information derived from the data gathered during sessions provided supervisors a good insight in patient progress and allowed for timely intervention where required.

Although, where possible, data input was structured and made quicker using drop-down lists and other multiple-choice widgets, some elements of the intervention required considerable free-text input. We provided keyboards for this purpose and though this may suit some users, others may find such input difficult on a tablet device.

Requirement 6: Training Environment

Although generally positive, the responses to items 3 and 7 on the SUS scale regarding EoU and learning rate, respectively, indicate that users were not always confident in using the system. Requests for an interface with *dummy* patients for each of the sessions (which was implemented during the pilot) support our belief that these shortcomings can be remedied with improved in-app training facilities.

Requirement 7: Automated Communication

During the pilot, automated communication was limited to patient-specific warnings being shown in the Web interfaces for supervisors and researchers. Such automated communications should be extended to automated emails to supervisors and, where necessary, to routine care health providers involved in the care of patients that make use of the system to fully comply with accountability requirements.

Requirement 8: User-Initiated Communication

Due to time constraints, user-initiated communications were not implemented during the pilot study, but CHWs were able to contact supervisors using the device's normal phone app and Google Hangouts. A WhatsApp group was used frequently for communication among CHWs and researchers involved in the trial. The limited implementation of this user-initiated communication in the pilot study version of the system does not seem to have had a negative effect on the good relationship between health care workers in a team, but this is likely because of the fact that these health care teams were already well established.

Requirement 9: Delivery in the Community

Mobile delivery of the intervention in patients' homes was facilitated using the mobile intervention interface on Android tablets that contained data for all the patients under the care of a given CHW using the device. Most tablets were shared by, normally, 2 CHWs, and the protocols required users to synchronize their devices before leaving the community health center. In a few instances, this sharing resulted in CHWs using a different tablet that did not contain the data for their patients. A login in the app with subsequent synchronization would have resulted in all data being downloaded, but this was not always performed, on occasion resulting in CHWs not being able to perform the session as planned. Although the tablets were equipped with subscriber identification module cards, a connection to the system's servers may not always be available at a patient's home. For this reason, it is desirable to implement a procedure within the app that unobtrusively ensures the data for the CHW using the tablet are indeed available on the device.

Task-shifting has been the preferred strategy to address the lack of specialized resources in LMICs. However, this problem is of such magnitude that even transferring responsibilities to low-cadre health workers is not enough to overcome the problem. More innovative solutions are needed. One such solution is for technology to provide assistance to these low-cadre health workers. This is already happening in many parts of the world, and the uptake is increasing rapidly [48]. Our pilot suggests that technology indeed may have a positive effect on the ability of low-cadre health workers to effectively deliver a psychosocial mental health intervention. Although our results lack statistical power, they suggest the intervention improved symptoms of depression and that the support provided to the CHWs by the system allowed them to deliver the intervention with very little training and support. Moreover, the pilot addresses the needs of a significantly underserved

population, which does not have access to routine mental health services delivered by highly skilled practitioners. The combination of task-shifting and technology is a key enabler for provision of mental health services to this underserved population.

Limitations

The timelines and resources available for the development of the system were considerably challenging and for this reason, choices had to be made in regard to the implementation of the requirements identified in the Methods section. Shortcomings in this regard were identified in the previous section and form the starting point for further development.

A further limitation related to the usability testing of our system concerns the limited number of users who provided their feedback (8 of the 11 users during the pilot). These numbers do not allow for a robust quantitative analysis of results obtained from the SUS and TAM questionnaires. Nevertheless, the qualitative analysis of their responses has provided invaluable feedback for further development of the system.

As a last limitation, as the supervisor role in the pilot was covered by psychologists involved in the development of the project, we have not yet robustly assessed their experiences with and opinion of the system. This will be rectified during the main trial.

Conclusions

In this paper, we proposed a set of high-level software requirements for ICT systems supporting task-shifted interventions based on the recommendation framework proposed by Campbell and Scott [8]. These software requirements were used to develop an ICT system for the support of a depression intervention provided by CHWs without specialized mental health training. The system consists of an app used by CHWs to provide a structured intervention to older adults in São Paulo's (Brazil) poorer neighborhoods and Web interfaces that allow the various stakeholders to monitor and supervise the intervention remotely. The intervention app was built using a purpose-built API that allows direct interaction between the graphical user interface and an underlying data source. The advantage of this approach is that new functionality can be rapidly added by defining the user interface alone. As a result, future development can be performed by users with limited experience in app development.

The SUS and TAM were used to elicit information on usability, PU, and EoU. Results from these assessments and insights gained during the pilot were used to assess the high-level software requirements we proposed. Our overall conclusion is that these high-level requirements adequately captured the functionality required to enable the CHWs to provide the intervention successfully. Nevertheless, the analysis of results indicated that some improvements are required for the system to be useable in a task-shifted intervention. The most important of these are more extensive access to a training environment, access for supervisors to metadata such as duration of sessions or exercises to identify issues, and a more robust and human-error–proof approach to availability of patient data on the mobile devices used during the intervention. Given the rapidly increasing number of task-shifted interventions in health care, which are delivered mostly by CHWs in LMICs, and other nonspecialized health workers in high-income countries, ICT solutions are a promising avenue to provide the support and accountability required in the performance of their demanding tasks.

Acknowledgments

This paper was written with the support of the Medical Research Council (the United Kingdom, ML/l016702/1) and the Fundação de Amparo à Pesquisa do Estado de São Paulo (Brazil, 2013/50953-4). The authors are thankful to the Associação Saúde da Família, Secretaria Municipal da Saúde de São Paulo, UBS Jardim Guarani, and UBS Vila Barbosa for the support provided in developing and testing this system. The authors are also thankful to DEBASÉ F+C-Brazil for their work on the media resources and to Dr. Luiz Antônio D'Angelo, from Associação Saúde da Família, for his support during all stages of this study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Questionnaires used in the qualitative study.
[PDF File (Adobe PDF File), 63 KB - jmir_v21i11e11346_app1.pdf]

References

1. World Health Organisation. Task Shifting: Global Recommendations and Guidelines. Geneva, Switzerland: World Health Organisation; 2008.
2. Zhu NS, Ling Z, Shen J, Lane J, Hu S. Factors associated with the decline of the Cooperative Medical System and barefoot doctors in rural China. Bull World Health Organ 1989;67(4):431-441 [FREE Full text] [doi: 10.1016/0168-8510(90)90372-k] [Medline: 2805221]
3. Sringernyuang L, Hongvivatana T, Pradabmuk P. World Health Organization. 1994. Community Health Workers and Drugs - A Case Study of Thailand. URL: http://apps.who.int/medicinedocs/en/d/Js2208e/ [accessed 2019-09-04]

4. Terry B, Bisanzo M, McNamara M, Dreifuss B, Chamberlain S, Nelson SW, et al. Task shifting: meeting the human resources needs for acute and emergency care in Africa. Afr J Emerg Med 2012 Dec;2(4):182-187. [doi: 10.1016/j.afjem.2012.06.005]
5. Maier CB, Aiken LH. Task-shifting from physicians to nurses in Europe and other major OECD countries. Eur J Public Health 2015 Oct;25(Suppl 3):ckv168.006. [doi: 10.1093/eurpub/ckv168.006]
6. Fulton BD, Scheffler RM, Sparkes SP, Auh EY, Vujicic M, Soucat A. Health workforce skill mix and task shifting in low income countries: a review of recent evidence. Hum Resour Health 2011 Jan 11;9:1 [FREE Full text] [doi: 10.1186/1478-4491-9-1] [Medline: 21223546]
7. Lehmann U, van Damme W, Barten F, Sanders D. Task shifting: the answer to the human resources crisis in Africa? Hum Resour Health 2009 Jun 21;7:49-50 [FREE Full text] [doi: 10.1186/1478-4491-7-49] [Medline: 19545398]
8. Campbell C, Scott K. Retreat from Alma Ata? The WHO's report on Task Shifting to community health workers for AIDS care in poor countries. Glob Public Health 2011;6(2):125-138. [doi: 10.1080/17441690903334232] [Medline: 19916089]
9. Bastawrous A, Rono HK, Livingstone IA, Weiss HA, Jordan S, Kuper H, et al. Development and validation of a smartphone-based visual acuity test (Peek Acuity) for clinical practice and community-based fieldwork. JAMA Ophthalmol 2015 Aug;133(8):930-937 [FREE Full text] [doi: 10.1001/jamaophthalmol.2015.1468] [Medline: 26022921]
10. Abas M, Bowers T, Manda E, Cooper S, Machando D, Verhey R, et al. 'Opening up the mind': problem-solving therapy delivered by female lay health workers to improve access to evidence-based care for depression and other common mental disorders through the Friendship Bench Project in Zimbabwe. Int J Ment Health Syst 2016;10:39 [FREE Full text] [doi: 10.1186/s13033-016-0071-9] [Medline: 27175215]
11. Philips M, Zachariah R, Venis S. Task shifting for antiretroviral treatment delivery in sub-Saharan Africa: not a panacea. Lancet 2008 Feb 23;371(9613):682-684. [doi: 10.1016/S0140-6736(08)60307-4] [Medline: 18295026]
12. Zachariah R, Ford N, Philips M, Lynch S, Massaquoi M, Janssens V, et al. Task shifting in HIV/AIDS: opportunities, challenges and proposed actions for sub-Saharan Africa. Trans R Soc Trop Med Hyg 2009 Jun;103(6):549-558. [doi: 10.1016/j.trstmh.2008.09.019] [Medline: 18992905]
13. Madhombiro M, Dube M, Dube-Marimbe B, Mutsvuke W. The challenges in implementing task shifting in mental health service in Zimbabwe. Ann Psychiatry Ment Health 2017;5(2):1-3 [FREE Full text]
14. World Health Organization. 2017. Depression and Other Common Mental Disorders: Global Health Estimates. URL: https://apps.who.int/iris/bitstream/handle/10665/254610/WHO-MSD-MER-2017.2-eng.pdf [accessed 2019-09-04]
15. Djernes JK. Prevalence and predictors of depression in populations of elderly: a review. Acta Psychiatr Scand 2006 May;113(5):372-387. [doi: 10.1111/j.1600-0447.2006.00770.x] [Medline: 16603029]
16. Alvarado BE, Zunzunegui MV, Béland F, Sicotte M, Tellechea L. Social and gender inequalities in depressive symptoms among urban older adults of Latin America and the Caribbean. J Gerontol B Psychol Sci Soc Sci 2007 Jul;62(4):S226-S236. [doi: 10.1093/geronb/62.4.s226] [Medline: 17673536]
17. Guerra M, Prina AM, Ferri CP, Acosta D, Gallardo S, Huang Y, et al. A comparative cross-cultural study of the prevalence of late life depression in low and middle income countries. J Affect Disord 2016 Jan 15;190:362-368 [FREE Full text] [doi: 10.1016/j.jad.2015.09.004] [Medline: 26544620]
18. da Silva SA, Scazufca M, Menezes PR. Population impact of depression on functional disability in elderly: results from 'São Paulo ágeing & health study' (SPAH). Eur Arch Psychiatry Clin Neurosci 2013 Mar;263(2):153-158. [doi: 10.1007/s00406-012-0345-4] [Medline: 22872105]
19. Huang H, Russo J, von Korff M, Ciechanowski P, Lin E, Ludman E, et al. The effect of changes in depressive symptoms on disability status in patients with diabetes. Psychosomatics 2012;53(1):21-29 [FREE Full text] [doi: 10.1016/j.psym.2011.07.001] [Medline: 22221718]
20. Gallo JJ, Bogner HR, Morales KH, Post EP, ten Have T, Bruce ML. Depression, cardiovascular disease, diabetes, and two-year mortality among older, primary-care patients. Am J Geriatr Psychiatry 2005 Sep;13(9):748-755 [FREE Full text] [doi: 10.1176/appi.ajgp.13.9.748] [Medline: 16166403]
21. Lino VT, Portela MC, Camacho LA, Atie S, Lima MJ. Assessment of social support and its association to depression, self-perceived health and chronic diseases in elderly individuals residing in an area of poverty and social vulnerability in Rio de Janeiro city, Brazil. PLoS One 2013;8(8):e71712 [FREE Full text] [doi: 10.1371/journal.pone.0071712] [Medline: 23951227]
22. Scazufca M, Menezes PR, Almeida OP. Caregiver burden in an elderly population with depression in São Paulo, Brazil. Soc Psychiatry Psychiatr Epidemiol 2002 Sep;37(9):416-422. [doi: 10.1007/s00127-002-0571-6] [Medline: 12242617]
23. Huang H, Menezes PR, da Silva SA, Tabb K, Barkil-Oteo A, Scazufca M. The association between depressive disorders and health care utilization: results from the São Paulo ageing and health study (SPAH). Gen Hosp Psychiatry 2014;36(2):199-202 [FREE Full text] [doi: 10.1016/j.genhosppsych.2013.11.003] [Medline: 24342113]
24. Katon WJ, Lin E, Russo J, Unutzer J. Increased medical costs of a population-based sample of depressed elderly patients. Arch Gen Psychiatry 2003 Sep;60(9):897-903. [doi: 10.1001/archpsyc.60.9.897] [Medline: 12963671]
25. Ciechanowski P, Wagner E, Schmaling K, Schwartz S, Williams B, Diehr P, et al. Community-integrated home-based depression treatment in older adults: a randomized controlled trial. J Am Med Assoc 2004 Apr 7;291(13):1569-1577. [doi: 10.1001/jama.291.13.1569] [Medline: 15069044]

26. Hunkeler EM, Katon W, Tang L, Williams Jr JW, Kroenke K, Lin EH, et al. Long term outcomes from the IMPACT randomised trial for depressed elderly patients in primary care. Br Med J 2006 Feb 4;332(7536):259-263 [FREE Full text] [doi: 10.1136/bmj.38683.710255.BE] [Medline: 16428253]
27. Unützer J, Katon W, Callahan CM, Williams Jr JW, Hunkeler E, Harpole L, IMPACT Investigators. Improving Mood-Promoting Access to Collaborative Treatment. Collaborative care management of late-life depression in the primary care setting: a randomized controlled trial. J Am Med Assoc 2002 Dec 11;288(22):2836-2845. [doi: 10.1001/jama.288.22.2836] [Medline: 12472325]
28. Charney DS, Reynolds CF, Lewis L, Lebowitz BD, Sunderland T, Alexopoulos GS, Depression and Bipolar Support Alliance. Depression and Bipolar Support Alliance consensus statement on the unmet needs in diagnosis and treatment of mood disorders in late life. Arch Gen Psychiatry 2003 Jul;60(7):664-672. [doi: 10.1001/archpsyc.60.7.664] [Medline: 12860770]
29. Ell K. Depression care for the elderly: reducing barriers to evidence-based practice. Home Health Care Serv Q 2006 Jun 20;25(1-2):115-148. [doi: 10.1300/j027v25n01_07] [Medline: 16803741]
30. Jacob K, Sharan P, Mirza I, Garrido-Cumbrera M, Seedat S, Mari JJ, et al. Mental health systems in countries: where are we now? Lancet 2007 Sep 22;370(9592):1061-1077. [doi: 10.1016/S0140-6736(07)61241-0] [Medline: 17804052]
31. Saxena S, Thornicroft G, Knapp M, Whiteford H. Resources for mental health: scarcity, inequity, and inefficiency. Lancet 2007 Sep 8;370(9590):878-889. [doi: 10.1016/S0140-6736(07)61239-2] [Medline: 17804062]
32. Saraceno B, van Ommeren M, Batniji R, Cohen A, Gureje O, Mahoney J, et al. Barriers to improvement of mental health services in low-income and middle-income countries. Lancet 2007 Sep 29;370(9593):1164-1174. [doi: 10.1016/S0140-6736(07)61263-X] [Medline: 17804061]
33. Macinko J, Guanais FC, de Fátima M, de Souza M. Evaluation of the impact of the Family Health Program on infant mortality in Brazil, 1990-2002. J Epidemiol Community Health 2006 Jan;60(1):13-19 [FREE Full text] [doi: 10.1136/jech.2005.038323] [Medline: 16361449]
34. Lehmann U, Friedman I, Sanders D. Semantic Scholar. 2004. Review of the Utilisation and Effectiveness of Community-Based Health Workers in Africa. URL: https://pdfs.semanticscholar.org/7a0c/fd44975b00b045c7957c9327a87c9fcc6b63.pdf [accessed 2019-09-06]
35. Haines A, Sanders D, Lehmann U, Rowe AK, Lawn JE, Jan S, et al. Achieving child survival goals: potential contribution of community health workers. Lancet 2007 Jun;369(9579):2121-2131. [doi: 10.1016/s0140-6736(07)60325-0] [Medline: 17586307]
36. Ekers DM, Webster L, van Straten A, Cuijpers P, Richards D, Gilbody S. Behavioural activation for depression; an update of meta-analysis of effectiveness and sub group analysis. PLoS One 2014 Jun 17;9(6):e100100. [doi: 10.1371/journal.pone.0100100] [Medline: 24936656]
37. Ekers D, Dawson MS, Bailey E. Dissemination of behavioural activation for depression to mental health nurses: training evaluation and benchmarked clinical outcomes. J Psychiatr Ment Health Nurs 2013 Mar;20(2):186-192. [doi: 10.1111/j.1365-2850.2012.01906.x] [Medline: 22452364]
38. Ekers D, Richards D, McMillan D, Bland JM, Gilbody S. Behavioural activation delivered by the non-specialist: phase II randomised controlled trial. Br J Psychiatry 2018 Jan 2;198(1):66-72. [doi: 10.1192/bjp.bp.110.079111] [Medline: 21200079]
39. Patel V, Weobong B, Weiss HA, Anand A, Bhat B, Katti B, et al. The healthy activity program (HAP), a lay counsellor-delivered brief psychological treatment for severe depression, in primary care in India: a randomised controlled trial. Lancet 2017 Jan;389(10065):176-185. [doi: 10.1016/s0140-6736(16)31589-6] [Medline: 27988143]
40. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613. [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
41. Scazufca M, de Paula Couto MCP, Henrique MG, Mendes AV, Matijasevich A, Pereda PC, et al. Pilot study of a two-arm non-randomized controlled cluster trial of a psychosocial intervention to improve late life depression in socioeconomically deprived areas of São Paulo, Brazil (PROACTIVE): feasibility study of a psychosocial intervention for late life depression in São Paulo. BMC Public Health 2019 Aug 22;19(1):1152 [FREE Full text] [doi: 10.1186/s12889-019-7495-5] [Medline: 31438903]
42. Scazufca M, Menezes P, Tabb K, Kester R, Rössler W, Huang H. Identification and treatment of depression of older adults in primary care: findings from the São Paulo ageing and health study. Fam Pract 2015 Jul 28;33(3):233-237. [doi: 10.1093/fampra/cmv062] [Medline: 26219991]
43. Lancet Global Mental Health Group, Chisholm D, Flisher AJ, Lund C, Patel, Patel S, et al. Scale up services for mental disorders: a call for action. Lancet 2007 Oct;370(9594):1241-1252. [doi: 10.1016/S0140-6736(07)61242-2] [Medline: 17804059]
44. Brooke J. SUS - a quick and dirty usability scale. In: Jordan PW, Thomas B, McClelland IL, Weerdmeester B, editors. Usability Evaluation In Industry. Boca Raton, Florida, United States: CRC Press; 1996:189-194.
45. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep;13(3):319-340. [doi: 10.2307/249008]
46. Sauro J, Lewis JR. Quantifying the User Experience: Practical Statistics for User Research. Netherlands: Elsevier; 2012.

47. Bangor A, Kortum P, Miller J. Determining what individual SUS scores mean: adding an adjective rating scale. J Usability Stud 2009;4(3):114-123 [FREE Full text]
48. Naslund JA, Aschbrenner KA, Araya R, Marsch LA, Unützer J, Patel V, et al. Digital technology for treating and preventing mental disorders in low-income and middle-income countries: a narrative review of the literature. Lancet Psychiatry 2017 Jun;4(6):486-500. [doi: 10.1016/s2215-0366(17)30096-2] [Medline: 28433615]

Abbreviations

API: application programming interface
CHW: community health worker
EoU: ease of use
ICT: information and communications technology
LMICs: low-middle-income countries
NAs: nurse assistants
PHQ-9: 9-item Patient Health Questionnaire
PU: perceived usefulness
SUS: System Usability Scale
TAM: Technology Acceptance Model

Edited by G Eysenbach; submitted 12.07.18; peer-reviewed by M Pineros, E Chiauzzi; comments to author 06.10.18; revised version received 30.11.18; accepted 29.04.19; published 12.11.19.

Please cite as:
Van de Ven P, Araya R, P de Paula Couto MC, Henrique MG, Meere D, Vilela Mendes A, Peters TJ, Seabra A, Franzin RM, Carvalho Pereda P, Scazufca M
Investigating Software Requirements for Systems Supporting Task-Shifted Interventions: Usability Study
J Med Internet Res 2019;21(11):e11346
URL: https://www.jmir.org/2019/11/e11346
doi:10.2196/11346
PMID:31714246

Original Paper

Cost Per Participant Recruited From Rural and Remote Areas Into a Smoking Cessation Trial Via Online or Traditional Strategies: Observational Study

Judith Byaruhanga[1,2], MPH; Flora Tzelepis[1,2,3], PhD; Christine Paul[1,3], PhD; John Wiggers[1,2,3], PhD; Emma Byrnes[1,3], BSocSc (Hons); Christophe Lecathelinais[2], DESS

[1]University of Newcastle, Callaghan, Australia

[2]Hunter New England Population Health, Wallsend, Australia

[3]Hunter Medical Research Institute, New Lambton Heights, Australia

Corresponding Author:
Judith Byaruhanga, MPH
University of Newcastle
University Drive
Callaghan, 2308
Australia
Phone: 61 2 4924 6454
Email: judith.byaruhanga@uon.edu.au

Abstract

Background: Rural and remote residents are more likely to smoke than those who live in major cities; however, recruitment of research participants from rural and remote areas can be challenging. The cost per participant recruited from rural and remote areas via online (eg, social media) and traditional strategies (eg, print) has implications for researchers on how to allocate resources to maximize the number of participants recruited. Participant characteristics such as demographics, financial stress, mental health, and smoking-related factors may be associated with recruitment method (ie, online vs traditional), and so it is important to understand whether certain subgroups are more likely to be recruited via a particular strategy.

Objective: This study aimed to determine the cost per participant recruited and examine whether characteristics such as demographics, financial stress, mental health, and smoking-related factors may be associated with the recruitment method (ie, online vs traditional).

Methods: Participants were recruited into a randomized trial that provided smoking cessation support. Eligible participants were aged 18 years or older; used tobacco daily; had access to video communication software, internet, and telephone; had an email address; and lived in a rural or remote area of New South Wales, Australia. This study describes the natural (observed) experience of recruiting participants via online and traditional methods into a smoking cessation trial.

Results: Over 17 months, 655 participants were recruited into the smoking cessation trial. A total of 88.7% (581/655) of the participants were recruited via online methods. Moreover, 1.8% (12/655) of the participants were recruited from remote locations and none from very remote areas. The cost per participant recruited by the various online strategies ranged from Aus $7.29 (US $4.96, £4.09, and €4.43) for Gumtree, a local online classified website, to Aus $128.67 (US $87.63, £72.20, and €78.28) for email. The cost per participant recruited using traditional strategies ranged from Aus $0 (US $0, £0, and €0) for word of mouth to Aus $3990.84 (US $2757.67, £2227.85, and €2477.11) for telephone. Women had greater odds of being recruited via online methods than men (odds ratio 2.50, 95% CI 1.42-4.40). No other characteristics were associated with the recruitment method.

Conclusions: The cost per participant recruited via online and traditional strategies varied, with the range being smaller for online than traditional recruitment strategies. Women have greater odds of being recruited via online strategies into rural smoking cessation trials.

Trial Registration: Australian New Zealand Clinical Trials Registry ACTRN12617000514303; https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=372584&isReview=true

(J Med Internet Res 2019;21(11):e14911) doi:10.2196/14911

KEYWORDS

smoking cessation; tobacco use; rural population

Introduction

Tobacco Use in Rural Population

In high-income countries such as Australia [1] and the United States [2], people who live in rural and remote areas have been identified as a priority population for smoking cessation interventions. Research has found that factors such as lower income [3], lower education levels [4], and fewer substance abuse treatment facilities [4,5] contribute to higher rates of tobacco use in rural areas, and rural residents are more likely to smoke tobacco than their counterparts who live in major cities [4]. Rural and remote locations comprise all areas outside major cities [1,4]. In the United States, the smoking prevalence is 24.9% in rural areas [6] compared with 20.8% of US adults in the general population [7]. In Australia, those in remote and very remote areas are 1.7 times more likely to smoke than those in major cities [8]. In Australia, the smoking rate is 12.8% [9] in the general population compared with 22% [10] among those living in regional and remote areas. There is, therefore, a need for tobacco control research to focus on such high-risk populations; however, recruitment of research participants who live in rural or remote locations for smoking cessation interventions can be challenging [11]. Several factors such as large geographical areas, limited resources, and transport barriers contribute to the difficulty of recruiting participants from these locations [12].

Recruitment Methods

Timely recruitment of participants is a very important aspect of large research studies [13] because delays in recruitment may increase financial costs [14], reduce the sample size obtained, and limit the study's robustness through inadequate statistical power [15,16], resulting in the reduced possibility of detecting a statistically significant result when there is a true difference between treatments [17]. Researchers conducting trials, therefore, often explore several avenues to recruit participants ranging from online to traditional methods. Traditional recruitment approaches that have been used include print (eg, direct mail, newspapers, posters, and flyers), broadcast advertising (eg, radio and television advertisements) [18], and word of mouth. Online methods such as social media (eg, Facebook and Twitter) [19,20], online advertisements (eg, Google advertisements), email, and other website promotions have also been used to recruit participants into research [21-23]. The potential reach of online recruitment strategies makes them appealing for recruiting participants into smoking cessation programs [23-25] and smoking cessation trials [26-29]. For example, in 2018, there were 3.9 billion internet users [30], and in 2016 to 2017, 86% of all households in Australia had access to the internet [31]. Specifically, 82.7% of those living in inner regional locations of Australia have internet access at home, as do 80.7% in outer regional locations and 77.1% in remote areas [31].

There is also evidence that online recruitment strategies are able to recruit participants with different demographic characteristics to those recruited via traditional recruitment methods [26-28,32,33]. Google advertisements, links on websites [27], and Facebook advertising are significantly more likely to recruit smokers aged between 18 and 25 years [32] or find a 7-year age difference between traditional and online recruitment [26]. In contrast, participants recruited by traditional methods such as newspapers, flyers, radio, and word of mouth in another smoking cessation trial were older and more confident in their quit attempts compared with those recruited via online methods [28]. Furthermore, people with depression have been found to be more likely to use the internet for health purposes [34,35], and social media users are more likely to have mental health conditions [36]. Financial stress, defined as "the difficulty that an individual or household may have in meeting basic financial commitments due to a shortage of money" [37], may also affect whether people have access to the internet and online advertisements; however, no research has examined whether financial stress is associated with recruitment strategy (ie, online vs traditional).

Cost of Recruitment

Traditional and online recruitment strategies have been found to be cost effective [26,28,29,38]. For example, in 2016, an Australian study reported that the cost of an enrolled participant into a smoking cessation trial was Aus $56.34 (US $38.93, £31.43, and €34.95) for online (social media) and Aus $52.33 (US $36.13, £29.20, and €32.46) for traditional recruitment strategies [28]. In another Australian smoking cessation trial, the cost of online social media recruitment was Aus $42.34 (US $29.24, £23.63, and €26.27) per participant enrolled and Aus $21.52 (US $14.86, £12.01, and €13.35) per participant for traditional recruitment strategies [26]. Furthermore, a US study with young adult smokers used Facebook and spent up to US $10 (£8.08 and €8.98) in 2014 for each participant recruited into the smoking cessation trial [38]. Similarly, in a study conducted in the United States using other online strategies but not Facebook (ie, an online health risk assessment [HRA], online advertisements, traditional material [offline promotions], and quit-line screening), young adult smokers were recruited for US $41.35 (£33.41 and €37.10) for each enrolled smoker [29]. Online materials recruited the most smokers (online advertisements: n=1426; US $41.35) for the lowest cost in comparison with the other 3 recruitment strategies (HRA: n=397, US $630.85; quit-line screening: n=189, US $133.61; and offline materials: n=1341, US $56.23) [29]. This cost information has also been presented as a table in Multimedia Appendix 1. The existing research on the cost of recruiting smokers into research trials and characteristics of participants recruited by online versus traditional methods has focused on the general population [26-29] and suggests that online recruitment is a feasible method that can complement traditional recruitment strategies [26,28].

Objectives

To our knowledge, no studies have compared the cost per participant recruited and subgroups of smokers from rural and remote locations recruited using online and traditional strategies

into a smoking cessation trial. Therefore, this study described the different methods used for recruitment of smokers from rural and remote areas and aimed to (1) determine the cost per participant recruited for each recruitment strategy and (2) examine whether demographic characteristics, financial stress, mental health, and smoking-related characteristics of the participants were associated with the type of recruitment method (ie, online vs traditional).

Methods

Design

This study used the baseline survey of a smoking cessation trial that assessed how participants were recruited into the study (ie, via online or traditional strategies). Participants were recruited to take part in a 3-arm randomized trial of behavioral support for smoking cessation [39]. The 3 arms of the trial provided (1) real-time video counseling delivered via video communication software (eg, Skype (Microsoft Corporation) and FaceTime (Apple Inc), (2) telephone counseling, or (3) written self-help materials (control) [39]. The University of Newcastle Human Research Ethics Committee granted ethics approval. The trial is prospectively registered with the Australian New Zealand Clinical Trials Registry (ACTRN12617000514303).

Participants

Participants were recruited over a 17-month period from May 25, 2017, to October 2, 2018. To be eligible for the trial, smokers needed to be aged 18 years or older; use tobacco daily; have access to a mode of video communication (eg, Skype and FaceTime), the internet, and a telephone; have a current email address; and reside in an inner or outer regional, remote, or very remote area of New South Wales (NSW) Australia. The classification of these areas was based on the Accessibility and Remoteness Index of Australia (ARIA+) [40]. The ARIA+ is a geographic accessibility index that reflects the ease or difficulty people face in accessing services in nonmetropolitan Australia based on residential postcode [40]. The different degrees of rurality are established on road distances needed to be traveled from the location to service centers of various population sizes. The resulting index of geographic accessibility classification is then used to categorize locations as inner regional (>0.2 to ≤2.4), outer regional (>2.4 to ≤5.92), remote (>5.92 to 10.53), and very remote (>10.53) areas [40].

Recruitment

Both traditional and online advertising methods were used concurrently to recruit participants. The success of the recruitment approaches was monitored on an ongoing basis, and recruitment resources were allocated to each method based on participant numbers recruited by each method throughout the recruitment period.

Online Recruitment Strategies

For all the online strategies, there were advertisements throughout the entire recruitment period of May 2017 to October 2018.

Facebook

All Facebook advertisements targeted only adults in rural and remote postcodes that were eligible for this study as classified by the ARIA+. A total of 41 paid advertisements were placed on Facebook as needed to boost the recruitment rate with a fixed daily budget limit of Aus $30 to Aus $100 (US $20.73-US $69.10). Unpaid Facebook promotions were posted on a daily basis to more than 1161 Facebook community boards.

Twitter

A total of 38 Twitter posts were made from the study twitter account and used to promote the research project using the hashtag quit smoking (#quitsmoking).

Emails to Businesses and Individuals

A total of 11,858 emails obtained from a business directory and partners in health services, such as local health districts, were sent out to corporate businesses, health organizations, mining companies, local area businesses, and trades people located in the eligible areas.

Gumtree

A total of 9 advertisements were placed on Gumtree, a local free online classified website, in postcodes that were eligible, and these remained on the website for the entire recruitment period.

Web Promotions and Internet Search

A study website was set up at the beginning of the study to promote the study and was active for the entire study period. The website was visible to anyone that searched online search engines for smoking cessation information and was cited on all print and online study advertising material.

Traditional Recruitment Strategies

Newspaper

The newspaper recruitment strategy was used from August 2017 to September 2018 with a total of 51 different rural and remote newspapers. Two sizes of paid advertisements were used (92×63 mm or 100×72 mm) and were placed in the middle section of the newspaper. There were 44 free newspaper articles and 7 paid newspaper advertisements in different rural and remote newspapers. The free articles were written by journalists together with the research team to promote the research.

Posters

In total, 68 Men's Sheds located in rural and remote NSW were telephoned in October 2017 and asked to display a poster promoting the study on their premises. Of these, 10 premises indicated that they were willing to do so, and a colored poster was distributed to them. In May 2017, posters were sent to 28 gyms, 21 Police-Citizens Youth clubs, 5 dental centers, 2 pharmacies, 2 hairdressers, 1 laundry shop, and 1 restaurant. Given the geographic diversity of locations, the study team did not visit these sites to verify if in fact the posters were displayed.

Flyers

A total of 2400 flyers were distributed. Flyers were printed in color on 1 side of an A4 paper and trifolded. A total of 19,00

flyers were placed into residential mailboxes in inner and outer regional locations of NSW in June, July, November, and December 2017, and 500 flyers were distributed to attendees at a rural conference in September 2018.

Radio

A total of 6 live radio interviews on different radio stations targeting rural and remote areas were conducted in July, August, September, and November 2017 and September 2018. The interviews promoted the smoking cessation trial, including information about the study and how to enroll into the study.

Telephone Calls

Telephone calls were made to 2465 households (2284 landlines and 181 mobiles) from an unlisted landline between June 2017 and September 2018. A single call was made to these households to determine if there was a potentially eligible participant and, if so, whether they would be interested in enrolling into the study.

Magazine

A single, colored advertisement, 136×61 mm in size, that promoted the study was placed in the middle section of a rural magazine in April 2017.

Data Collection Procedures

All the advertising materials promoted the study website. Potential participants could click on (for online) or type (for traditional) the study website address to access information about the smoking cessation trial. The study website included a detailed information letter that described the research, outlined the importance of the study, informed who was eligible to participate, and contained a hyperlink that potential participants could select to complete the online screening survey. The screening survey computer software determined whether eligibility requirements were met, and if participants were eligible, they were immediately redirected to the online baseline survey. The baseline survey took about 10 to 15 min to complete.

Measures

Recruitment Method

To determine whether participants were recruited via an online or traditional strategy, each participant was asked how they heard about the study during the baseline survey. The response options were as follows: Facebook advertisement or page, Twitter, Gumtree, Google advertisements, Web promotions (eg, study promoted on websites), Instagram, email, posters, flyers, postcards, magazine, newspaper, television, radio, word of mouth, telephone, or other (please specify).

Sociodemographic Characteristics

The baseline survey assessed sociodemographic characteristics including age, gender, country of birth, Aboriginal or Torres Strait Islander origin, education, marital status, and occupational status.

Geographical Location

The participant's residential postcode collected during the baseline survey was used to determine location of residence (ie, inner regional, outer regional, remote, or very remote) according to the ARIA+ [40]. ARIA+ is a geographic accessibility index that reflects the ease or difficulty people face in accessing services. It is based on road distances needed to be traveled from the location to service centers of various population sizes from a point to the nearest urban centers and localities in 5 separate population ranges [40]. The classification comprises major cities, inner regional Australia, outer regional Australia, remote Australia, and very remote Australia [40].

Mental Health

Anxiety and depression were assessed using the 4-item Patient Health Questionnaire-4 (PHQ-4). The PHQ-4 is a reliable and valid measure of depression and anxiety in the general population [41]. The participants were asked, over the last 2 weeks, how often they had been bothered by the following: (1) feeling nervous, anxious, or on edge; (2) not being able to stop or control worrying, (3) little interest or pleasure in doing things, and (4) feeling down, depressed, or hopeless. The response options were as follows: not at all, several days, more than half the days, or nearly every day.

Financial Stress

Financial stress was assessed by asking if in the past 6 months, the participant did any of the following because of a shortage of money: (1) could not pay electricity, gas, or telephone bills on time; (2) could not pay the mortgage or rent on time; (3) pawned or sold something; (4) went without meals; (5) was unable to heat home; (6) asked for financial help from friends or family; and (7) asked for help from a welfare or community organization. The response options were yes or no [42].

Quitting Intentions

Participants were asked what best described their intentions regarding quitting. The response options were as follows: will quit in the next 30 days, will quit in the next 6 months, may quit in the future but not in the next 6 months, never expect to quit, and do not know.

Nicotine Dependence

The Heaviness of Smoking Index (HSI), which comprises 2 questions (time to first cigarette of day and number of cigarettes per day), was used to assess nicotine dependence [43]. The HSI was used rather than the Fagerstrom Test for Nicotine Dependence [44] because it has fewer items and it has been found to have acceptable reliability and validity [45].

Confidence to Quit

Confidence to quit was assessed by asking "if you decided to give up smoking altogether, how likely do you think you would be to succeed?" The response options were as follows: (1) very likely, (2) fairly likely, (3) fairly unlikely, (4) very unlikely, and (5) do not know.

Statistical Analysis

Statistical analyses were completed using SAS software version 9.3 (SAS Institute Inc). Categorical measures were described

using frequencies and percentages. For continuous measures, means, medians, and standard deviations were reported. To assess the relationship between variables of interest and recruitment strategy, a multiple logistic regression model was run, which included the recruitment strategy as the dependent outcome and the following as independent variables: sociodemographic characteristics, mental health measures, financial stress, and smoking-related variables. Odds ratios (ORs) and 95% CIs were reported. A criterion for statistical significance of $P<.05$ was used.

Cost Analysis

All monetary costs were recorded directly from invoices as they were incurred. Costs associated with each recruitment strategy were tracked and summarized. The traditional recruitment costs included staff time for telephone calls; preparing materials; delivering radio interviews; liaising with newspaper journalists; dropping materials into mailboxes; printing and postage of materials; and costs for paid newspaper advertisements, magazine advertisements, and resources (eg, printing materials). The costs associated with the online strategies included staff time to send emails; posting advertisements on Facebook, Twitter, and Gumtree; setting up the study website, and the cost of paid Facebook advertisements. The cost per participant recruited was calculated by summing the total cost for each recruitment strategy and dividing it by the number of participants recruited via that strategy.

Results

Participant Characteristics

Between May 25, 2017, and October 2, 2018, 655 participants were enrolled in the study. Moreover, 77.4% (507/655) of the participants were female, 87.0% (570/655) were Australian born, and 40.3% (264/655) had completed a certificate or diploma from a Technical and Further Education campus, which provided vocational education in Australia. The average age of the participants was 43.7 years, 55.4% (363/655) participants were married or living in a de facto relationship, 62.6% (410/655) were employed, and 73.0% (477/655) lived in an inner regional area. The mean number of cigarettes smoked per day was 19.0 (SD 9.2), and the average minutes to first cigarette after waking up was 37.3 (SD 86.5) min. The mean age participants started smoking regularly was 16.6 (SD 4.5) years. Almost half of the participants (48.6%, 318/655) had made a quit attempt in the last 12 months, 45.0% (295/655) intended to quit in the next 30 days, and 53.0% (347/655) were confident that they were likely to quit (Table 1).

Table 1. Participants' characteristics (N=655).

Demographic characteristics	Value
Age (years), mean (SD); median	43.7 (11.8); 43.0
Gender, n (%)	
Female	507 (77.4)
Male	148 (22.6)
Country of birth, n (%)	
Australia	570 (87.0)
Other	85 (13.0)
Aboriginal or Torres Strait Islander, n (%)	
Yes	53 (8.1)
No or do not know	602 (91.9)
Education, n (%)	
Completed primary school	5 (0.8)
Completed years 7-9	52 (7.9)
Completed year 10	127 (19.4)
Completed Higher Secondary Certificate or leaving year 12	52 (7.9)
Technical and Further Education Certificate or diploma	264 (40.3)
University degree or higher	146 (22.2)
Other	9 (1.4)
Marital status, n (%)	
Never married	147 (22.4)
Married or de facto	363 (55.4)
Married but separated	46 (7.0)
Divorced	74 (11.3)
Widowed	16 (2.4)
Don't know	9 (1.4)
Location, n (%)	
Inner regional	477 (73.0)
Outer regional	164 (25.1)
Remote	12 (1.8)
Employment, n (%)	
Full time or part time or casual	410 (62.6)
Home duties	73 (11.2)
Permanently unable to work or ill	56 (8.6)
Retired	35 (5.3)
Student (full or part time)	27 (4.1)
Unemployed	40 (6.1)
Other	14 (2.1)
Smoking characteristics	
Cigarettes per day (n=620), mean (SD); median	19.0 (9.21); 20
Time to first cigarette (n=632), mean (SD); median	37.3 (86.5); 15
Age (years) started smoking regularly (n=649), mean (SD); median	16.6 (4.5); 16
Quit attempt in the past 12 months, n (%)	

Demographic characteristics	Value
Yes	318 (48.6)
No	335 (51.2)
Don't know	1 (0.2)
Quitting intentions, n (%)	
Will quit in the next 30 days	295 (45.0)
Will quit in the next 6 months	229 (35.0)
Will not quit in the next 6 months or never expect to quit or do not know	131 (20.0)
Confidence to quit, n (%)	
Very or fairly likely to quit	347 (53.0)
Very or fairly unlikely to quit	168 (25.7)
Do not know	140 (21.4)

Recruitment Strategies

Table 2 outlines the percentage of participants recruited by online and traditional methods. A total of 88.7% (581/655) of the participants were recruited via online strategies and 11.3% (74/655) of the participants were via traditional strategies. The majority (83.5%, 547/655) of the participants reported that they heard about the study through Facebook, whereas 7.2% (47/655) were recruited via newspaper.

Cost per Smoker Recruited

The cost per participant recruited for each online and traditional recruitment strategy are described in Table 2. The cost for each participant recruited by the different online strategies ranged from Aus $7.29 for Gumtree to Aus $128.67 for email. In contrast, for the traditional recruitment strategies, the cost per participant recruited ranged from Aus $0 for word of mouth and television (initiated and completed by television station at no cost to the study) to Aus $3990.84 for telephone recruitment.

Comparison of Participant Characteristics Recruited via Online Versus Traditional Strategies

Table 3 compares the characteristics of participants recruited via online or traditional methods. Women had greater odds of being recruited via online strategies than men (OR 2.50, 95% CI 1.42-4.40; *P*=.002). There were no significant associations between any other participant characteristics and type of recruitment method.

Table 2. Participant recruitment via recruitment methods.

Recruitment method	Participants, n (%)	Total cost for each strategy				Cost per participant recruited			
		Aus $	US $	£	€	Aus $	US $	£	€
Online methods									
Gumtree	5 (0.8)	36.43	24.81	20.44	22.16	7.29	4.96	4.09	4.43
Web promotions and internet search	10 (1.5)	437.56	298.00	245.55	266.22	43.76	29.80	24.56	26.62
Twitter	1 (0.2)	61.52	41.90	34.52	37.43	61.52	41.90	34.52	37.43
Facebook	547 (83.5)	33,738.52	22,977.28	18,932.47	20,526.24	61.68	42.01	34.61	37.53
Email	18 (2.7)	2315.98	1577.32	1299.56	1409.07	128.67	87.63	72.20	78.28
Traditional methods									
Word of mouth	19 (2.9)	0	0	0	0	0	0	0	0
Television (initiated or completed by television station—no cost to study)	1 (0.2)	0	0	0	0	0	0	0	0
Newspaper	47 (7.2)	2363.38	1609.60	1326.27	1437.90	50.28	34.25	28.22	30.59
Radio (live interviews)	2 (0.3)	205.55	139.99	115.36	125.06	102.78	70.00	57.68	62.53
Magazine	2 (0.3)	170.81	116.33	95.86	103.92	85.41	51.96	47.93	58.17
Posters	1 (0.2)	566.65	385.92	318.01	344.76	566.65	385.92	318.01	344.76
Flyers	1 (0.2)	2546.29	1734.18	1429.38	1549.64	2546.29	1734.18	1429.38	1549.64
Telephone	1 (0.2)	3990.84	2757.67	2227.85	2477.11	3990.84	2757.67	2227.85	2477.11

Table 3. Comparison of participants recruited via online and traditional strategies.

Variable and categories	Online strategies (n=581)	Traditional strategies (n=74)	Odds ratio (95% CI)	*P* value
Gender, n (%)				.002
Female	460 (90.7)	47 (9.3)	2.50 (1.42-4.40)	
Male	121 (81.8)	27 (18.2)	Reference	
Age (years) mean (SD)	43.72 (11.7)	43.23 (12.8)	1.01 (0.98-1.03)	.52
Education, n (%)				.39
Year 10 or less	166 (89.7)	19 (10.3)	1.56 (0.75-3.26)	
Higher Secondary Certificate or year 12 or Technical and Further Education	290 (90.1)	32 (9.9)	1.50 (0.79-2.83)	
University or tertiary	124 (84.9)	22 (15.1)	Reference	
Marital status, n (%)				.09
With partner	328 (90.4)	35 (9.6)	1.61 (0.94-2.77)	
Without partner	253(86.6)	39 (13.4)	Reference	
Employment, n (%)				.21
Employed full or casual or part time	366 (89.3)	44(10.7)	1.45 (0.81-2.59)	
Not employed	215 (87.8)	30 (12.2)	Reference	
Aboriginal or Torres Strait Islander, n (%)				.14
Yes	49 (92.5)	4 (7.6)	2.55 (0.73-8.87)	
No	532 (88.4)	70 (11.6)	Reference	
Australian born, n (%)				.15
No	79 (92.9)	6 (7.1)	1.95 (0.79-4.81)	
Yes	502 (88.1)	68 (11.9)	Reference	
Anxiety, n (%)				.87
No	295 (88.3)	39 (11.7)	1.05 (0.57-1.95)	
Yes	286 (89.1)	35 (10.9)	Reference	
Depression, n (%)				.41
No	348 (88.3)	46 (11.7)	0.76 (0.40-1.46)	
Yes	233 (89.3)	28 (10.7)	Reference	
Financial stress, n (%)				.73
0	268 (88.5)	35 (11.6)	0.71 (0.29-1.72)	
1-3	241 (89.3)	29 (10.7)	0.82 (0.34-1.96)	
4-7	72 (87.8)	10 (12.2)	Reference	
Quitting intentions, n (%)				.31
Will quit in the next 30 days	267 (90.5)	28 (9.5)	1.68 (0.86-3.28)	
Will quit in the next 6 months	203 (88.7)	26 (11.4)	1.41 (0.71-2.80)	
Will not quit within the 6 months, never expect to quit, and do not know	111 (84.7)	20 (15.3)	Reference	
Confidence to quit, n (%)				.40
Very or fairly likely	313 (90.2)	34 (9.8)	1.25 (0.74-2.11)	
Very or fairly unlikely or do not know	268 (87.0)	40 (13.0)	Reference	
Heaviness Smoking Index score, n (%)				.57
0-2: low addiction	143 (87.2)	21 (12.8)	0.60 (0.23-1.58)	
3-4: moderate addiction	307 (88.0)	42 (12.0)	0.65 (0.27-1.58)	

Variable and categories	Online strategies (n=581)	Traditional strategies (n=74)	Odds ratio (95% CI)	*P* value
5-6: high addiction	79 (90.8)	8 (9.2)	Reference	

Discussion

Principal Findings

This Australian study found that both online and traditional recruitment strategies could facilitate the recruitment of smokers who lived in rural and remote areas into a smoking cessation trial. The online strategy had a high yield of smokers residing in rural and remote locations, similar to other studies conducted with the wider general population [26,29,32,38,46-50]. This is the first study to compare the cost per participant recruited from rural and remote locations using online and traditional strategies into a smoking cessation trial. The findings provide insight on the variability of cost per participant depending on the recruitment strategy used. The cost per participant recruited ranged from Aus $7.29 to Aus $128.67 for online strategies and Aus $0 to Aus $3990.84 for traditional strategies. Word of mouth was, by far, the most inexpensive method per participant recruited because the study incurred no cost from this recruitment strategy. However, word of mouth would not have been sufficient to recruit enough participants into the smoking cessation trial and, therefore, is not practical on its own, although it is an inexpensive strategy.

Facebook recruited the largest number of participants in this study compared with other online strategies and traditional recruitment methods (although this would have been influenced by the most money being spent on Facebook). However, Facebook was neither the most inexpensive online recruitment method in terms of cost per participant recruited nor was it cheaper than every traditional recruitment strategy. Nevertheless, to maximize the number of rural and remote participants recruited into smoking cessation trials, there needs to be a balance between the cost per participant recruited and the capacity (ie, total participants' that can access each method). Specifically, there would be substantially fewer participants recruited into the study if only the most inexpensive methods (eg, word of mouth and Gumtree) were used, as these strategies did not recruit as large a number of participants as Facebook.

Comparison With Prior Studies

The cost per participant for recruiting rural and remote smokers is variable and comparable with recruiting smokers in the general population [26,28,29,32,38,49,50], especially given that this study included all the costs such as staff time and printing of flyers unlike other studies [26,28,29,38,49]. For instance, in a US study of a Web-based smoking intervention [49], the total advertisement cost per randomized participant was US $40.51 (£33.33 and €36.32) for Facebook, US $34.71 (£28.56 and €31.12) for Google, and US $20.30 (£16.7 and €18.2) for traditional sources but excluded the costs for personnel. In another Australian study, the cost of online social media recruitment was Aus $42.34 (US $29.24, £23.63, and €26.27) per participant enrolled in a smoking cessation trial, whereas for the traditional recruitment strategies, it was Aus $21.52 (US $14.86, £12.01, and €13.35) per participant [26]. However, these costs did not include call back costs for after-hours' time and materials sometimes mailed out to the newspaper participants [26]. In 2016, another Australian study [28] reported that the costs to recruit an eligible respondent into a smoking cessation trial via online (social media) methods was Aus $57.34 (US $38.93, £31.43, and €34.95) versus Aus $52.33 (US $36.13, £29.20, and €32.46) for the traditional methods. However, some costs such as printing and distribution of flyers were not included, and radio interviews were conducted free of charge [28]. Another smoking cessation trial [29] reported that online advertisements cost US $41.35 (£33.41 and €37.10), whereas traditional materials cost US $56.23 per enrolled participant. Finally, another study of young adult smokers using Facebook recruitment reported a cost of US $8.80 per eligible participant recruited into a cessation trial [38].

The results from this study showed that compared with men, women had greater odds of being recruited via online strategies. This is similar to a study by Stanczyk et al [50] where the internet yielded a higher proportion of female smokers, compared with other strategies. However other smoking cessation trials [26,28,51] have found no association between gender and method of recruitment. Nonetheless, it should be noted that the majority of men in this study were still recruited via online methods. Bearing in mind that most participants recruited via online strategies were recruited via Facebook, the finding that women had greater odds of being recruited via online strategies might be explained by evidence that women, unlike men, considered Facebook an integral part of their life to connect daily [52] and were more likely to seek support, either social (Facebook) or otherwise [53], but this was not the case for men and, particularly, men living in rural areas [54]. The participation of fewer men in smoking cessation research compared with women is not surprising. For instance, in a systematic review of proactive telephone counseling for smoking cessation [55], 17 out of the 21 studies reported a greater proportion of women than men participated [55]. Similarly, in a smoking cessation trial conducted in rural Kansas, United States, only 35% of the sample were men [56].

There were no significant associations found for other characteristics between participants recruited via online and traditional strategies in this study. In contrast, a study by Frandsen et al [28] reported a significant association between social media recruitment and younger participants compared with those recruited by traditional methods into smoking cessation support. Furthermore, although there is evidence that social media users are more likely to have mental health conditions [36], this study did not find any association between anxiety or depression and the recruitment method.

Limitations

The study had a number of limitations. First, access to the internet was an eligibility criterion and might have excluded some smokers who would have been interested in behavioral

support and had no internet connection. Second, most people were recruited from inner or outer regional areas, and the recruitment strategies were only able to reach a small proportion of smokers from remote areas and no one from very remote areas. This may not be surprising given that few newspapers and online notice boards or local classified websites specifically targeted very remote areas. In addition, the study targeted only 0.001% of the NSW population that lives in very remote areas [57]. Third, another limitation is that self-report was used as the measure of how participants were recruited into the study. Although there are online systems to track users after having clicked on an online advertisement (eg, via Facebook), people who heard about the study from traditional sources (eg, hardcopy newspaper) could not be tracked via such a system. Finally, this study recruited only adults living in rural and remote areas of Australia, and therefore, the findings might have limited generalizability to other countries, particularly low- and middle-income countries.

Conclusions

The cost per participant recruited via online and traditional strategies varied with the range of costs being smaller for online than traditional recruitment strategies. Women have greater odds of being recruited via online strategies into rural smoking cessation trials.

Acknowledgments

This study and author FT were supported by a Cancer Institute New South Wales Early Career Fellowship (15/ECF/1-44). CP was supported by a National Health and Medical Research Council Career Development Fellowship (1061335). Hunter New England Population Health and the Hunter Medical Research Institute provided infrastructure support. The sponsors did not have any role in the study design; collection, analysis, and interpretation of data; writing of the paper; and decision to submit the paper for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary table of studies reporting cost per participant recruited in the general population via online and traditional recruitment strategies (summarises literature in Introduction).
[PDF File (Adobe PDF File), 128 KB - jmir_v21i11e14911_app1.pdf]

References

1. Australian Institute of Health and Welfare. 2016. Australia's Health 2016 URL: https://www.aihw.gov.au/getmedia/9844cefb-7745-4dd8-9ee2-f4d1c3d6a727/19787-AH16.pdf.aspx [accessed 2019-09-18]
2. Doogan NJ, Roberts ME, Wewers ME, Stanton CA, Keith DR, Gaalema DE, et al. A growing geographic disparity: rural and urban cigarette smoking trends in the United States. Prev Med 2017 Nov;104:79-85 [FREE Full text] [doi: 10.1016/j.ypmed.2017.03.011] [Medline: 28315761]
3. vander Weg MW, Cunningham CL, Howren MB, Cai X. Tobacco use and exposure in rural areas: findings from the behavioral risk factor surveillance system. Addict Behav 2011 Mar;36(3):231-236. [doi: 10.1016/j.addbeh.2010.11.005] [Medline: 21146318]
4. Eberhardt MS, Freid VM, Harper S, Ingram DD, Makuc DM, Pamuk E, et al. Centers for Disease Control and Prevention. 2001. Health, United States, 2001: Urban and Rural Health Chartbook URL: https://www.cdc.gov/nchs/data/hus/hus01cht.pdf [accessed 2019-09-18]
5. Eberhardt MS, Pamuk ER. The importance of place of residence: examining health in rural and nonrural areas. Am J Public Health 2004 Oct;94(10):1682-1686. [doi: 10.2105/ajph.94.10.1682] [Medline: 15451731]
6. Doescher MP, Jackson JE, Jerant A, Hart LG. Prevalence and trends in smoking: a national rural study. J Rural Health 2006;22(2):112-118. [doi: 10.1111/j.1748-0361.2006.00018.x] [Medline: 16606421]
7. Pleis JR, Lethbridge-Cejku M. Summary health statistics for US adults: national health interview survey, 2005. Vital Health Stat 10 2006 Dec(232):1-153 [FREE Full text] [doi: 10.1037/e403882008-001] [Medline: 17252928]
8. National Rural Health Alliance. 2014. Smoking and Rural Health: Fact Sheet URL: http://www.ruralhealth.org.au/sites/default/files/publications/nrha-factsheet-smoking.pdf [accessed 2019-12-03]
9. Australian Institute of Health and Welfare. Drug statistics series #28, Cat. #PHE 183. National Drug Strategy Household Survey detailed report 2013. Canberra: Australian Institute of Health and Welfare; 2014.
10. Australian Institute of Health and Welfare. 2014. Australia's Health 2014 URL: https://www.aihw.gov.au/getmedia/d2946c3e-9b94-413c-898c-aa5219903b8c/16507.pdf.aspx?inline=true [accessed 2019-09-18]
11. Morgan LL, Fahs PS, Klesh J. Barriers to research participation identified by rural people. J Agric Saf Health 2005 Nov;11(4):407-414. [doi: 10.13031/2013.19719] [Medline: 16381161]

12. Friedman DB, Bergeron CD, Foster C, Tanner A, Kim SH. What do people really know and think about clinical trials? A comparison of rural and urban communities in the south. J Community Health 2013 Aug;38(4):642-651. [doi: 10.1007/s10900-013-9659-z] [Medline: 23468319]
13. Patrick JH, Pruchno RA, Rose MS. Recruiting research participants: a comparison of the costs and effectiveness of five recruitment strategies. Gerontologist 1998 Jun;38(3):295-302. [doi: 10.1093/geront/38.3.295] [Medline: 9640849]
14. Treweek S, Wilkie E, Craigie AM, Caswell S, Thompson J, Steele RJ, BeWEL Team. Meeting the challenges of recruitment to multicentre, community-based, lifestyle-change trials: a case study of the BeWEL trial. Trials 2013 Dec 18;14:436 [FREE Full text] [doi: 10.1186/1745-6215-14-436] [Medline: 24351063]
15. Sully BG, Julious SA, Nicholl J. A reinvestigation of recruitment to randomised, controlled, multicenter trials: a review of trials funded by two UK funding agencies. Trials 2013 Jun 9;14:166 [FREE Full text] [doi: 10.1186/1745-6215-14-166] [Medline: 23758961]
16. Relton C, Torgerson D, O'Cathain A, Nicholl J. Rethinking pragmatic randomised controlled trials: introducing the 'cohort multiple randomised controlled trial' design. Br Med J 2010 Mar 19;340:c1066. [doi: 10.1136/bmj.c1066] [Medline: 20304934]
17. Carlisle B, Kimmelman J, Ramsay T, MacKinnon N. Unsuccessful trial accrual and human subjects protections: an empirical analysis of recently closed trials. Clin Trials 2015 Feb;12(1):77-83 [FREE Full text] [doi: 10.1177/1740774514558307] [Medline: 25475878]
18. Rait MA, Prochaska JJ, Rubinstein ML. Recruitment of adolescents for a smoking study: use of traditional strategies and social media. Transl Behav Med 2015 Sep;5(3):254-259 [FREE Full text] [doi: 10.1007/s13142-015-0312-5] [Medline: 26327930]
19. Thornton LK, Harris K, Baker AL, Johnson M, Kay-Lambkin FJ. Recruiting for addiction research via Facebook. Drug Alcohol Rev 2016 Jul;35(4):494-502. [doi: 10.1111/dar.12305] [Medline: 26179224]
20. O'Connor A, Jackson L, Goldsmith L, Skirton H. Can I get a retweet please? Health research recruitment and the Twittersphere. J Adv Nurs 2014 Mar;70(3):599-609. [doi: 10.1111/jan.12222] [Medline: 23909740]
21. Pedersen ER, Helmuth ED, Marshall GN, Schell TL, PunKay M, Kurz J. Using Facebook to recruit young adult veterans: online mental health research. JMIR Res Protoc 2015 Jun 1;4(2):e63 [FREE Full text] [doi: 10.2196/resprot.3996] [Medline: 26033209]
22. Valdez RS, Guterbock TM, Thompson MJ, Reilly JD, Menefee HK, Bennici MS, et al. Beyond traditional advertisements: leveraging Facebook's social structures for research recruitment. J Med Internet Res 2014 Oct 27;16(10):e243 [FREE Full text] [doi: 10.2196/jmir.3786] [Medline: 25348050]
23. Muñoz RF, Barrera AZ, Delucchi K, Penilla C, Torres LD, Pérez-Stable EJ. International Spanish/English internet smoking cessation trial yields 20% abstinence rates at 1 year. Nicotine Tob Res 2009 Sep;11(9):1025-1034 [FREE Full text] [doi: 10.1093/ntr/ntp090] [Medline: 19640833]
24. Graham AL, Fang Y, Moreno JL, Streiff SL, Villegas J, Muñoz RF, et al. Online advertising to reach and recruit Latino smokers to an internet cessation program: impact and costs. J Med Internet Res 2012 Aug 27;14(4):e116 [FREE Full text] [doi: 10.2196/jmir.2162] [Medline: 22954502]
25. Barrera AZ, Pérez-Stable EJ, Delucchi KL, Muñoz RF. Global reach of an internet smoking cessation intervention among Spanish- and English-speaking smokers from 157 countries. Int J Environ Res Public Health 2009 Mar;6(3):927-940 [FREE Full text] [doi: 10.3390/ijerph6030927] [Medline: 19440423]
26. Frandsen M, Walters J, Ferguson SG. Exploring the viability of using online social media advertising as a recruitment method for smoking cessation clinical trials. Nicotine Tob Res 2014 Feb;16(2):247-251. [doi: 10.1093/ntr/ntt157] [Medline: 24127266]
27. Gordon JS, Akers L, Severson HH, Danaher BG, Boles SM. Successful participant recruitment strategies for an online smokeless tobacco cessation program. Nicotine Tob Res 2006 Dec;8(Suppl 1):S35-S41. [doi: 10.1080/14622200601039014] [Medline: 17491169]
28. Frandsen M, Thow M, Ferguson SG. The effectiveness of social media (Facebook) compared with more traditional advertising methods for recruiting eligible participants to health research studies: a randomized, controlled clinical trial. JMIR Res Protoc 2016 Aug 10;5(3):e161 [FREE Full text] [doi: 10.2196/resprot.5747] [Medline: 27511829]
29. Buller DB, Meenan R, Severson H, Halperin A, Edwards E, Magnusson B. Comparison of 4 recruiting strategies in a smoking cessation trial. Am J Health Behav 2012 Sep;36(5):577-588 [FREE Full text] [doi: 10.5993/AJHB.36.5.1] [Medline: 22584086]
30. Statista. 2019. Number of Internet Users Worldwide From 2005 to 2018 (in Millions) URL: https://www.statista.com/statistics/273018/number-of-internet-users-worldwide/ [accessed 2019-12-03]
31. Australian Bureau of Statistics, Australian Government. 2018. 8146.0 - Household Use of Information Technology, Australia, 2016-17 URL: http://www.abs.gov.au/Ausstats/abs%40.nsf/mf/8146.0 [accessed 2019-09-18]
32. Ramo DE, Hall SM, Prochaska JJ. Reaching young adult smokers through the internet: comparison of three recruitment mechanisms. Nicotine Tob Res 2010 Jul;12(7):768-775 [FREE Full text] [doi: 10.1093/ntr/ntq086] [Medline: 20530194]

33. Nolte MT, Shauver MJ, Chung KC. Analysis of four recruitment methods for obtaining normative data through a web-based questionnaire: a pilot study. Hand (N Y) 2015 Sep;10(3):529-534 [FREE Full text] [doi: 10.1007/s11552-014-9730-y] [Medline: 26330790]
34. Bessière K, Pressman S, Kiesler S, Kraut R. Effects of internet use on health and depression: a longitudinal study. J Med Internet Res 2010 Feb 28;12(1):e6 [FREE Full text] [doi: 10.2196/jmir.1149] [Medline: 20228047]
35. Park S, Lee SW, Kwak J, Cha M, Jeong B. Activities on Facebook reveal the depressive state of users. J Med Internet Res 2013 Oct 1;15(10):e217 [FREE Full text] [doi: 10.2196/jmir.2718] [Medline: 24084314]
36. Pantic I. Online social networking and mental health. Cyberpsychol Behav Soc Netw 2014 Oct;17(10):652-657 [FREE Full text] [doi: 10.1089/cyber.2014.0070] [Medline: 25192305]
37. Nicol A. Melbourne Institute. 2010. The Experience of Financial Stress in Australia: the Role of Life Events and Prior Stress URL: https://melbourneinstitute.unimelb.edu.au/assets/documents/hilda-bibliography/conference-papers-lectures/2010/Nicol_AIFS2010.pdf [accessed 2019-09-18]
38. Ramo DE, Rodriguez TM, Chavez K, Sommer MJ, Prochaska JJ. Facebook recruitment of young adult smokers for a cessation trial: methods, metrics, and lessons learned. Internet Interv 2014 Apr;1(2):58-64 [FREE Full text] [doi: 10.1016/j.invent.2014.05.001] [Medline: 25045624]
39. Tzelepis F, Wiggers J, Paul CL, Byaruhanga J, Byrnes E, Bowman J, et al. A randomised trial of real-time video counselling for smoking cessation in regional and remote locations: study protocol. Contemp Clin Trials 2018 Nov;74:70-75. [doi: 10.1016/j.cct.2018.10.001] [Medline: 30290277]
40. Australian Bureau of Statistics, Australian Government. 2018. Australian Statistical Geography Standard (ASGS): Volume 5 - Remoteness Structure, July 2016 URL: http://www.abs.gov.au/AUSSTATS/abs@.nsf/Latestproducts/1270.0.55.005Main%20Features15July%202016 [accessed 2019-09-18]
41. Löwe B, Wahl I, Rose M, Spitzer C, Glaesmer H, Wingenfeld K, et al. A 4-item measure of depression and anxiety: validation and standardization of the Patient Health questionnaire-4 (PHQ-4) in the general population. J Affect Disord 2010 Apr;122(1-2):86-95. [doi: 10.1016/j.jad.2009.06.019] [Medline: 19616305]
42. Siahpush M, Borland R, Scollo M. Smoking and financial stress. Tob Control 2003 Mar;12(1):60-66 [FREE Full text] [doi: 10.1136/tc.12.1.60] [Medline: 12612364]
43. Kozlowski LT, Porter CQ, Orleans CT, Pope MA, Heatherton T. Predicting smoking cessation with self-reported measures of nicotine dependence: FTQ, FTND, and HSI. Drug Alcohol Depend 1994 Feb;34(3):211-216. [doi: 10.1016/0376-8716(94)90158-9] [Medline: 8033758]
44. Heatherton TF, Kozlowski LT, Frecker RC, Fagerström KO. The Fagerström test for nicotine dependence: a revision of the Fagerström tolerance questionnaire. Br J Addict 1991 Sep;86(9):1119-1127. [doi: 10.1111/j.1360-0443.1991.tb01879.x] [Medline: 1932883]
45. Borland R, Yong HH, O'Connor RJ, Hyland A, Thompson ME. The reliability and predictive validity of the heaviness of smoking index and its two components: findings from the International Tobacco Control Four Country Study. Nicotine Tob Res 2010 Oct(12 Suppl):S45-S50 [FREE Full text] [doi: 10.1093/ntr/ntq038] [Medline: 20889480]
46. McDonnell D, Lee H, Kazinets G, Moskowitz JM. Online recruitment of targeted populations: lessons learned from a smoking cessation study among Korean Americans. Soc Mar Q 2010 Aug 26;16(3):2-22 [FREE Full text] [doi: 10.1080/15245004.2010.500441]
47. Danaher BG, McKay HG, Seeley JR. The information architecture of behavior change websites. J Med Internet Res 2005 May 18;7(2):e12 [FREE Full text] [doi: 10.2196/jmir.7.2.e12] [Medline: 15914459]
48. Lane TS, Armin J, Gordon JS. Online recruitment methods for web-based and mobile health studies: a review of the literature. J Med Internet Res 2015 Jul 22;17(7):e183 [FREE Full text] [doi: 10.2196/jmir.4359] [Medline: 26202991]
49. Watson NL, Mull KE, Heffner JL, McClure JB, Bricker JB. Participant recruitment and retention in remote ehealth intervention trials: methods and lessons learned from a large randomized controlled trial of two web-based smoking interventions. J Med Internet Res 2018 Aug 24;20(8):e10351 [FREE Full text] [doi: 10.2196/10351] [Medline: 30143479]
50. Stanczyk NE, Bolman C, Smit ES, Candel MJ, Muris JW, de Vries H. How to encourage smokers to participate in web-based computer-tailored smoking cessation programs: a comparison of different recruitment strategies. Health Educ Res 2014 Feb;29(1):23-40. [doi: 10.1093/her/cyt104] [Medline: 24287402]
51. Coday M, Richey P, Thomas F, Tran QT, Terrell SB, Tylavsky F, et al. The recruitment experience of a randomized clinical trial to aid young adult smokers to stop smoking without weight gain with interactive technology. Contemp Clin Trials Commun 2016 Apr 15;2:61-68 [FREE Full text] [doi: 10.1016/j.conctc.2015.12.010] [Medline: 26949747]
52. Thompson SH, Lougheed E. Frazzled by Facebook? An exploratory study of gender differences in social network communication among undergraduate men and women. Coll Stud J 2012;46(1):88-98 [FREE Full text]
53. Joiner R, Cuprinskaite J, Dapkeviciute L, Johnson H, Gavin J, Brosnan M. Gender differences in response to Facebook status updates from same and opposite gender friends. Comput Hum Behav 2016 May;58:407-412. [doi: 10.1016/j.chb.2016.01.008]
54. Roy P, Tremblay G, Robertson S. Help-seeking among male farmers: connecting masculinities and mental health. Sociol Ruralis 2014 Apr 16;54(4):460-476 [FREE Full text] [doi: 10.1111/soru.12045]

55. Tzelepis F, Paul CL, Walsh RA, McElduff P, Knight J. Proactive telephone counseling for smoking cessation: meta-analyses by recruitment channel and methodological quality. J Natl Cancer Inst 2011 Jun 22;103(12):922-941. [doi: 10.1093/jnci/djr169] [Medline: 21666098]
56. Richter K, Shireman TI, Ellerbeck EF, Cupertino AP, Catley D, Cox LS, et al. Comparative and cost effectiveness of telemedicine versus telephone counseling for smoking cessation. J Med Internet Res 2015 May 8;17(5):e113 [FREE Full text] [doi: 10.2196/jmir.3975] [Medline: 25956257]
57. Australian Bureau of Statistics, Australian Government. 2017. Regional Population Growth, Australia, 2014-15 URL: http://www.abs.gov.au/ausstats/abs@.nsf/Previousproducts/3218.0Main%20Features152014-15?opendocument&tabname=Summary&prodno=3218.0&issue=2014-15&num=&view [accessed 2019-08-18]

Abbreviations

ARIA+: Accessibility and Remoteness Index of Australia
HSI: Heaviness of Smoking Index
HRA: health risk assessment
NSW: New South Wales
OR: odds ratio
PHQ-4: Patient Health Questionnaire-4

Edited by G Eysenbach; submitted 03.06.19; peer-reviewed by N Walker, N Machado; comments to author 06.07.19; revised version received 09.08.19; accepted 29.08.19; published 12.11.19.

Please cite as:
Byaruhanga J, Tzelepis F, Paul C, Wiggers J, Byrnes E, Lecathelinais C
Cost Per Participant Recruited From Rural and Remote Areas Into a Smoking Cessation Trial Via Online or Traditional Strategies: Observational Study
J Med Internet Res 2019;21(11):e14911
URL: https://www.jmir.org/2019/11/e14911
doi:10.2196/14911
PMID:31714253

Original Paper

Cost and Effectiveness of Using Facebook Advertising to Recruit Young Women for Research: PREFER (Contraceptive Preferences Study) Experience

Edwina McCarthy[1], BA SocSci (Hons); Danielle Mazza[1], MBBS, MD, FRACGP, DRANZCOG, GAICD

Department of General Practice, School of Primary and Allied Health Care, Monash University, Melbourne, Australia

Corresponding Author:
Danielle Mazza, MBBS, MD, FRACGP, DRANZCOG, GAICD
Department of General Practice
School of Primary and Allied Health Care
Monash University
Building 1, 270 Ferntree Gully Rd
Notting Hill
Melbourne, 3168
Australia
Phone: 61 399024512
Fax: 61 399024300
Email: Danielle.mazza@monash.edu

Abstract

Background: Social media is a popular and convenient method for communicating on the Web. The most commonly used social networking website, Facebook, is increasingly being used as a tool for recruiting research participants because of its large user base and its ability to target advertisements on the basis of Facebook users' information.

Objective: We evaluated the cost and effectiveness of using Facebook to recruit young women into a Web-based intervention study (PREFER). The PREFER study aimed to determine whether an educational video could increase preference for and uptake of long-acting reversible contraception (LARC).

Methods: We placed an advertisement on Facebook over a 19-day period from December 2017 to January 2018, inviting 16- to 25-year-old women from Australia to participate in a Web-based study about contraception. Those who clicked on the advertisement were directed to project information, and their eligibility was determined by using a screening survey.

Results: Our Facebook advertisement delivered 130,129 impressions, resulting in over 2000 clicks at an overall cost of Aus $918 (Aus $0.44 per click). Web-based project information was accessed by 493 women. Of these, 462 women completed the screening survey, and 437 (437/463, 95%) women were eligible. A total of 322 young women participated in Surveys 1 and 2 (74% response rate), and 284 women participated in Survey 3 (88% retention rate), with an advertising cost of Aus $2.85 per consenting participant.

Conclusions: Facebook proved to be a quick, effective, and cost-efficient tool for recruiting young Australian women into a study that was investigating contraceptive preferences. However, Web-based recruitment may result in sociodemographic biases. Further research is required to evaluate whether Facebook is suitable for recruiting older study populations.

(J Med Internet Res 2019;21(11):e15869) doi:10.2196/15869

KEYWORDS

social media; Facebook; recruitment; intervention study; patient education; internet

Introduction

Participant recruitment for research is challenging. This process involves identifying potentially eligible individuals, implementing strategies to target potential participants, inviting participants into the study, and obtaining informed consent. Traditional recruitment strategies include flyers, newspaper advertisements, mail-outs, word of mouth, and television broadcasts. However, these methods are often slow, labor intensive, and expensive, and these can lead to project delays, and, in some cases, failure to meet recruitment targets. Furthermore, difficulties sourcing participants and their contact

details, the need to involve third-party organizations, the costs of staffing, travel, and printing, and the delay between obtaining consent and participation add further complexity.

Social media is a popular and convenient platform for communicating on the Web. Social media enables users to share information, such as updates, images, videos, and events, as well as send messages and maintain contact with other users. Facebook is the most commonly used social networking website, with over 2 billion users worldwide [1]. This includes 15 million users in Australia, with 50% of the Australian population using Facebook daily [2]. In Australia, over 2.5 million Facebook users are aged 18 to 25 years [2]. Facebook is increasingly being used as a tool for recruiting research participants. Its large user base and its ability to target advertisements according to demographic information made available by Facebook users make Facebook an effective recruitment approach. Facebook has been used successfully to recruit participants for studies on potentially sensitive topics, such as smoking cessation [3], alcohol consumption [4], abortion [5], and sexual health/HIV [6]. Facebook has also been successfully used to recruit participants for intervention studies on depression [7], posttraumatic stress disorder [8], physical activity [9], and smoking [10]. An Australian study that examined contraceptive use (CUPID) successfully used Facebook, as well as traditional methods, to recruit young women [11].

Although the use of Facebook as a recruitment technique is gaining popularity, the ability of researchers to draw upon the experiences of previous work is compromised by inconsistent reporting of Facebook recruitment. Thornton et al's [12] systematic review of studies that used Facebook to recruit participants found that only half of the articles described their Facebook recruitment strategy in detail. Comparing Facebook recruitment outcomes is further complicated by the use of different measures to assess effectiveness (eg, impressions and cost per click). Although the cost per participant is generally a comparable measure across different methods, few studies report on recruitment costs [13]. There is also limited information on the development of advertising content used in Facebook recruitment and the role of incentives with online recruitment.

We used Facebook advertising exclusively to target young Australian women and recruit them into a Web-based intervention study, the PREFER study. The PREFER study was concerned with examining whether an educational video could increase young women's preference for and uptake of LARC. This paper aimed to focus on our evaluation of the cost and effectiveness of using Facebook to recruit study participants.

Methods

Study Design

PREFER used a before-and-after survey methodology to address its aim. Participation in the study involved completing a series of Web-based surveys: (1) screening, (2) preintervention (Survey 1), (3) immediately postintervention (Survey 2), and (4) 6 months postintervention (Survey 3). This study was approved by the Monash University Human Research Ethics Committee (project number 10456).

Recruitment

We created a paid Facebook advertisement, featuring an image of young people using computer devices and a call for volunteers to participate in Web-based surveys about their contraceptive preferences. The Facebook Ads Manager fed the advertisement into newsfeeds, targeting women aged 16 to 25 years, who were living in Australia.

Women were eligible to participate in the survey if they (1) were aged between 16 and 25 years, (2) had been sexually active with a male partner in the past 6 months or anticipated sexual activity in the next 6 months, (3) had not undergone a tubal ligation or hysterectomy, (4) had a partner or partners who had not undergone a vasectomy, and (5) were not pregnant or had no desire to become pregnant in the next year.

Power and sample size estimations determined that 281 participants would be required in this study. The budget for the recruitment of participants was Aus $2000.

Data Collection

Facebook users who clicked on the advertisement were then taken to a project landing page via the Department of General Practice, Monash University website. The page outlined the phases of the study and included a link to the project's explanatory statement. Interested individuals were then able to access the Web-based screening survey and confirm that they had read and understood the explanatory statement. Eligible participants could then access the baseline survey (Survey 1), and the completion and submission of Web-based survey responses were considered implied consent. On completion of the baseline survey, participants were directed to view the long-acting reversible contraception (LARC) first Web-based patient education video on the Web (approximately 10 min long), which conveyed information on all contraceptive options available to Australian women. This included their mode of action, effectiveness, and side effects, starting with discussion of the LARC options with emphasis on their superior efficacy and patient acceptability. Following the video (intervention), participants were directed to complete Survey 2. A total of 6 months later, participants were emailed a Web-based link to the postintervention survey (Survey 3). Participants received up to Aus $40 in electronic gift vouchers as reimbursement for their time—Aus $20 for completing Surveys 1 and 2 and Aus $20 for completing Survey 3.

Data Analysis

We used the measures outlined in the following section to evaluate the cost and effectiveness of using Facebook to recruit female participants aged 16 to 25 years into the PREFER study.

Measures

Textboxes 1 and 2 show effectiveness and cost, respectively.

Textbox 1. Effectiveness measures.

- Impressions: The number of times the Facebook advertisement appeared on the Newsfeeds of Facebook users.
- Clicks: The number of times Facebook users clicked on the advertisement.
- Number of completed screening surveys.
- Eligibility: The proportion of interested individuals who were eligible to participate.
- Participants: The number of participants who completed Surveys 1 and 2.
- Retention rate: The proportion of participants who completed Surveys 1 and 2, who completed Survey 3.

Timeframe: Surveys 1 and 2 completed by participants within 5 months.

Textbox 2. Cost measures.

- Cost per click: The cost paid each time the advertisement was clicked during the advertising campaign.
- Cost per participant: The cost per consenting participant.
- Facebook advertising budget: Aus $2000 (excluding incentive payments of Aus $40 electronic gift card per participant).

Results

Overview

Our Facebook advertisement delivered 130,129 impressions, resulting in over 2000 clicks at an overall cost of Aus $918 (Aus $0.44/click). Web-based project information was accessed by 493 women. Of these, 462 women completed the screening survey and 437 (95%) women were eligible. This occurred in only 19 days. A total of 322 young women participated in Surveys 1 and 2 (74% response rate), and 284 women participated in Survey 3 (88% retention rate), with an advertising cost of Aus $2.85 per consenting participant. Recruitment outcomes are further described below.

Recruitment Outcomes

Textboxes 3 and 4 show effectiveness and cost, respectively.

Textbox 3. Effectiveness measures.

- Impressions: 130,129
- Clicks: 2101
- Completed screening survey: 462
- Eligibility: 95% (n=437)
- Participants: n=322
- Retention rate: 88% (n=284)

Timeframe: 19 days

Textbox 4. Cost measures.

- Cost per click: Aus $0.44
- Cost per participant: Aus $2.85
- Total Facebook advertising cost: Aus $918

Discussion

Principal Findings

Our study showed that Facebook advertising was a cost-effective method for recruiting participants into an educational intervention study delivered on the Web. Our project budget of Aus $2000 for a 5-month recruitment timeframe was met with an overall spend of Aus $918 over a brief 19-day Facebook advertising campaign. We recruited 322 young Australian women, of whom 284 (88%) women completed the follow-up survey (Survey 3). These results add to the evidence that Facebook is a useful tool for recruiting research participants [14,15].

Our study is novel in that we used a single paid Facebook advertisement to successfully recruit young women. Facebook recruitment proved to be an effective strategy for recruiting participants who were eligible. Our Facebook advertisement targeted users by age (16-25 years), sex (female), and location (Australia), and the screening survey was completed by 462 people. Importantly, this method captured individuals who were eligible and willing to participate (n=437; 95%), with the

remaining 5% ineligible because of not meeting the inclusion criteria. This confirms the findings of previous studies that showed that Facebook is effective in targeting and recruiting young Australian women [11,14,16]. However, unlike these studies, we were able to meet and exceed our recruitment target without combining Facebook recruitment with traditional recruitment methods or using multiple Facebook advertisements.

A surprising finding was that we achieved our participant target (n=281) within a 19-day timeframe. Owing to the cost-effectiveness of our recruitment campaign, we were able to increase our participant target from 281 to 320. Whitaker's [15] review showed that it took an average of 5 months (median 3 months, IQR 8) to recruit 463 participants (median 264 participants, IQR 775). Our study had anticipated a 5-month timeframe for the recruitment of participants for Surveys 1 and 2; however, recruitment was completed within 19 days. Over this period, our advertisement appeared in Facebook newsfeeds more than 130,000 times. Despite reports of poor participant retention in other internet-based studies [17], participant retention was high at our 6-month follow-up (88% completed Survey 3). This highlights the potential strengths of recruiting young women for research by using Facebook and conducting follow-up via email. Using the internet to connect with and communicate with research participants enables them respond and complete Web-based surveys at a time and place convenient for them.

Facebook recruitment was also an affordable approach for recruiting participants into our study. We used less than half (Aus $918) of our recruitment budget (Aus $2000), and the cost per participant was only Aus $2.85 (US $2.05). Traditional recruitment methods have been estimated to cost US $1094.27 for television, US $811.99 for print media, US $635.92 for radio, and US $332.46 for postal recruitment [18]. Previous studies resorted to combining Facebook advertising with traditional recruitment methods such as mail-outs, radio promotion, newspaper advertising, and word of mouth [11,19,20] to achieve recruitment targets. Other studies combined Facebook advertising with other Web-based methods, such as Google, Web-based newsletters and email [3,20], or other social media platforms, such as Twitter [6] and MySpace [21]. Our streamlined recruitment approach of developing 1 Facebook advertisement enabled us to meet and exceed our original recruitment target, without the need to apply other recruitment strategies or develop alternative Web-based advertising content as required by other studies [11,14]. This facilitated our ability to track the interest in our advertisement over the course of our recruitment campaign.

Our cost per participant was also inexpensive compared with other studies that used Facebook advertising for recruitment. A systematic review of studies that used Facebook to recruit participants for health research found that the average cost per click was US $0.51 [14], compared with our study's cost per click of US $0.32. Furthermore, our cost per participant was considerably less expensive than other studies that used Facebook to recruit young Australian women [11,14]. Our findings add to the growing evidence of the cost benefits of using Facebook to recruit research participants [12,15].

Limitations

A limitation to our study is that did not collect reasons for nonparticipation because of the quick progress of participant recruitment and data collection. Reasons for nonparticipation may have been particularly useful to know if our recruitment target was not met, as it can assist researchers to direct their resources more appropriately to help progress recruitment efforts [11].

Conclusions

For the purposes of recruiting young Australian women to our Web-based intervention, Facebook proved to be an affordable, effective, and quick method of recruiting participants. Our study adds important data on the outcomes of using Facebook to recruit research participants, and the study highlights the importance of documenting successful recruitment campaigns, particularly for sexual health research, which can be considered a sensitive topic. Our findings will inform researchers of the benefits of using Facebook to recruit participants; however, further research is required to establish the effectiveness of using Facebook to target other study populations, such as people in older age groups or from diverse cultural backgrounds.

Acknowledgments

This study was funded by Bayer Australia Ltd. Bayer did not contribute to the study design, data collection and analysis, the drafting of the manuscript, or the decision to submit the manuscript for publication, but they were offered an opportunity to comment on the manuscript before submission. The authors would like to acknowledge the young women who took the time to participate in the study. The authors would also like to acknowledge the team from the Office of Strategic Marketing and Communications, Monash University, for their assistance in the development of recruitment advertising content for this study.

Conflicts of Interest

None declared.

References

1. Facebook. URL: http://www.facebook.com [accessed 2019-07-01]
2. Social Media News Blog Australia. 2019 Aug 1. Social Media Statistics Australia – July 2019. URL: https://www.socialmedianews.com.au/social-media-statistics-australia-july-2019/ [accessed 2019-08-14]

3. Ramo DE, Hall SM, Prochaska JJ. Reaching young adult smokers through the internet: comparison of three recruitment mechanisms. Nicotine Tob Res 2010 Jul;12(7):768-775 [FREE Full text] [doi: 10.1093/ntr/ntq086] [Medline: 20530194]
4. Pedersen ER, Naranjo D, Marshall GN. Recruitment and retention of young adult veteran drinkers using Facebook. PLoS One 2017;12(3):e0172972 [FREE Full text] [doi: 10.1371/journal.pone.0172972] [Medline: 28249027]
5. Altshuler AL, Storey HL, Prager SW. Exploring abortion attitudes of US adolescents and young adults using social media. Contraception 2015 Mar;91(3):226-233. [doi: 10.1016/j.contraception.2014.11.009] [Medline: 25537853]
6. Yuan P, Bare MG, Johnson MO, Saberi P. Using online social media for recruitment of human immunodeficiency virus-positive participants: a cross-sectional survey. J Med Internet Res 2014 May 1;16(5):e117 [FREE Full text] [doi: 10.2196/jmir.3229] [Medline: 24784982]
7. Morgan AJ, Jorm AF, Mackinnon AJ. Internet-based recruitment to a depression prevention intervention: lessons from the Mood Memos study. J Med Internet Res 2013 Mar 12;15(2):e31 [FREE Full text] [doi: 10.2196/jmir.2262] [Medline: 23403043]
8. Brief DJ, Rubin A, Keane TM, Enggasser JL, Roy M, Helmuth E, et al. Web intervention for OEF/OIF veterans with problem drinking and PTSD symptoms: a randomized clinical trial. J Consult Clin Psychol 2013 Oct;81(5):890-900. [doi: 10.1037/a0033697] [Medline: 23875821]
9. Alley S, Jennings C, Plotnikoff RC, Vandelanotte C. An evaluation of web- and print-based methods to attract people to a physical activity intervention. JMIR Res Protoc 2016 May 27;5(2):e94 [FREE Full text] [doi: 10.2196/resprot.4826] [Medline: 27235075]
10. Watson NL, Mull KE, Heffner JL, McClure JB, Bricker JB. Participant recruitment and retention in remote eHealth intervention trials: methods and lessons learned from a large randomized controlled trial of two web-based smoking interventions. J Med Internet Res 2018 Aug 24;20(8):e10351 [FREE Full text] [doi: 10.2196/10351] [Medline: 30143479]
11. Harris ML, Loxton D, Wigginton B, Lucke JC. Recruiting online: lessons from a longitudinal survey of contraception and pregnancy intentions of young Australian women. Am J Epidemiol 2015 May 15;181(10):737-746. [doi: 10.1093/aje/kwv006] [Medline: 25883155]
12. Thornton L, Batterham PJ, Fassnacht DB, Kay-Lambkin F, Calear AL, Hunt S. Recruiting for health, medical or psychosocial research using Facebook: systematic review. Internet Interv 2016 May;4:72-81 [FREE Full text] [doi: 10.1016/j.invent.2016.02.001] [Medline: 30135792]
13. Amon KL, Campbell AJ, Hawke C, Steinbeck K. Facebook as a recruitment tool for adolescent health research: a systematic review. Acad Pediatr 2014;14(5):439-47.e4. [doi: 10.1016/j.acap.2014.05.049] [Medline: 25169155]
14. Fenner Y, Garland SM, Moore EE, Jayasinghe Y, Fletcher A, Tabrizi SN, et al. Web-based recruiting for health research using a social networking site: an exploratory study. J Med Internet Res 2012 Mar 1;14(1):e20 [FREE Full text] [doi: 10.2196/jmir.1978] [Medline: 22297093]
15. Whitaker C, Stevelink S, Fear N. The use of Facebook in recruiting participants for health research purposes: a systematic review. J Med Internet Res 2017 Aug 28;19(8):e290 [FREE Full text] [doi: 10.2196/jmir.7071] [Medline: 28851679]
16. Gunasekaran B, Jayasinghe Y, Brotherton JM, Fenner Y, Moore EE, Wark JD, et al. Asking about human papillomavirus vaccination and the usefulness of registry validation: a study of young women recruited using Facebook. Vaccine 2015 Mar 4;33(6):826-831. [doi: 10.1016/j.vaccine.2014.11.002] [Medline: 25444782]
17. Koo M, Skinner H. Challenges of internet recruitment: a case study with disappointing results. J Med Internet Res 2005 Mar 19;7(1):e6 [FREE Full text] [doi: 10.2196/jmir.7.1.e6] [Medline: 15829478]
18. Tate DF, LaRose JG, Griffin LP, Erickson KE, Robichaud EF, Perdue L, et al. Recruitment of young adults into a randomized controlled trial of weight gain prevention: message development, methods, and cost. Trials 2014 Aug 16;15:326 [FREE Full text] [doi: 10.1186/1745-6215-15-326] [Medline: 25128185]
19. Frandsen M, Thow M, Ferguson SG. The effectiveness of social media (Facebook) compared with more traditional advertising methods for recruiting eligible participants to health research studies: a randomized, controlled clinical trial. JMIR Res Protoc 2016 Aug 10;5(3):e161 [FREE Full text] [doi: 10.2196/resprot.5747] [Medline: 27511829]
20. Juraschek SP, Plante TB, Charleston J, Miller ER, Yeh H, Appel LJ, et al. Use of online recruitment strategies in a randomized trial of cancer survivors. Clin Trials 2018 Apr;15(2):130-138 [FREE Full text] [doi: 10.1177/1740774517745829] [Medline: 29361843]
21. Bull SS, Levine D, Schmiege S, Santelli J. Recruitment and retention of youth for research using social media: experiences from the Just/Us study. Vulnerable Child Youth Stud 2013;8(2):171-181. [doi: 10.1080/17450128.2012.748238]

Abbreviations

LARC: long-acting reversible contraception

Edited by G Eysenbach; submitted 14.08.19; peer-reviewed by A Cyr, E Da Silva; comments to author 05.10.19; revised version received 17.10.19; accepted 20.10.19; published 29.11.19.

Please cite as:
McCarthy E, Mazza D
Cost and Effectiveness of Using Facebook Advertising to Recruit Young Women for Research: PREFER (Contraceptive Preferences Study) Experience
J Med Internet Res 2019;21(11):e15869
URL: http://www.jmir.org/2019/11/e15869/
doi:10.2196/15869
PMID:31782738

Original Paper

Automatically Appraising the Credibility of Vaccine-Related Web Pages Shared on Social Media: A Twitter Surveillance Study

Zubair Shah[1,2], PhD; Didi Surian[1], PhD; Amalie Dyda[1], PhD; Enrico Coiera[1], PhD, MBBS; Kenneth D Mandl[3,4], MD, MPH; Adam G Dunn[1,4], PhD

[1]Centre for Health Informatics, Australian Institute of Health Innovation, Macquarie University, Sydney, Australia

[2]Division of Information and Communication Technology, College of Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar

[3]Department of Biomedical Informatics, Harvard Medical School, Boston, MA, United States

[4]Computational Health Informatics Program, Boston Children's Hospital, Boston, MA, United States

Corresponding Author:
Adam G Dunn, PhD
Centre for Health Informatics
Australian Institute of Health Innovation
Macquarie University
Sydney, 2109
Australia
Phone: 61 9850 2400
Email: adam.dunn@mq.edu.au

Abstract

Background: Tools used to appraise the credibility of health information are time-consuming to apply and require context-specific expertise, limiting their use for quickly identifying and mitigating the spread of misinformation as it emerges.

Objective: The aim of this study was to estimate the proportion of vaccine-related Twitter posts linked to Web pages of low credibility and measure the potential reach of those posts.

Methods: Sampling from 143,003 unique vaccine-related Web pages shared on Twitter between January 2017 and March 2018, we used a 7-point checklist adapted from validated tools and guidelines to manually appraise the credibility of 474 Web pages. These were used to train several classifiers (random forests, support vector machines, and recurrent neural networks) using the text from a Web page to predict whether the information satisfies each of the 7 criteria. Estimating the credibility of all other Web pages, we used the follower network to estimate potential exposures relative to a credibility score defined by the 7-point checklist.

Results: The best-performing classifiers were able to distinguish between low, medium, and high credibility with an accuracy of 78% and labeled low-credibility Web pages with a precision of over 96%. Across the set of unique Web pages, 11.86% (16,961 of 143,003) were estimated as low credibility and they generated 9.34% (1.64 billion of 17.6 billion) of potential exposures. The 100 most popular links to low credibility Web pages were each potentially seen by an estimated 2 million to 80 million Twitter users globally.

Conclusions: The results indicate that although a small minority of low-credibility Web pages reach a large audience, low-credibility Web pages tend to reach fewer users than other Web pages overall and are more commonly shared within certain subpopulations. An automatic credibility appraisal tool may be useful for finding communities of users at higher risk of exposure to low-credibility vaccine communications.

(J Med Internet Res 2019;21(11):e14007) doi:10.2196/14007

KEYWORDS

health misinformation; credibility appraisal; machine learning; social media

Introduction

Background

The spread of misinformation, which we define here to include communications that are not a fair representation of available evidence or communicate that evidence poorly, has become an increasingly studied topic in various domains [1-8]. Misinformation can cause harm by influencing attitudes and beliefs [9,10]. Although the rapid growth of Web-based communications has benefited public health by providing access to a much broader range of health information, most people trust health information available on the Web without attempting to validate the sources [11,12], despite concerns about the presence of misinformation in what they access [13] and known issues where biases and marketing can lead to the miscommunication of evidence [14-18]. Proposed approaches for mitigating the impact of misinformation include empowering individuals to better deal with the information they encounter and improvements in the automatic detection of misinformation on Web-based platforms [1].

Most studies aimed at finding or tracking misinformation on social media define misinformation using *veracity*—whether a claim is true or false or real or fake. In the health domain, veracity alone often does not provide enough information to be useful in understanding the range of factors that might influence attitudes and behaviors, such as persuasiveness, timeliness, or applicability. The *credibility* of health communications thus includes a broader set of factors that include veracity as well as readability and clarity, the use and transparency of sources, biases and false balance, and disclosure of conflicts of interest [19]. It is important to consider credibility when evaluating the potential impact of health communications on health attitudes and outcomes because certain types of communication can be true but misleading, such as in the case of false balance in news media [20].

A range of tools have been developed to assess the credibility of health information available on the Web. Most were designed as checklists to be used by experts to assess the credibility and transparency of what they are reading. The DISCERN tool was designed as a general purpose tool for evaluating the quality of health information [21], with an emphasis on Web pages that patients might use to support the decisions they make about their health. The Quality Index for health-related Media Reports (QIMR) is a more recent example and differs from previous tools in that it was designed to be used to evaluate the quality of communications about new biomedical research [22]. Common elements of the tools used by experts to assess the credibility of health research reporting and patient information on the Web include the following: the veracity of the included information, transparency about sources of evidence, disclosure of advertising, simplicity and readability of the language, and use of balanced language that does not distort or sensationalize [19]. Most of the tools can be time-consuming to use and often require specific training or expertise to apply. Organizations such as HealthNewsReview that ended in 2018 used experts to evaluate new health-related communications as they appear in the news media [23].

Public perception of vaccines is an exemplar of the problem of misinformation spread through news and social media [24]. Beyond public health and vaccines, previous studies using social media data derived from Twitter to understand the spread and impact of misinformation have variously extracted text from what users post or information about their social connections [25-29]. Attitudes toward vaccines and opinions about disease outbreaks are a common application domain studied in social media research [30-34]. In particular, studies of human papillomavirus (HPV) vaccines have made use of the information users post and their social connections, as well as what people might have been exposed to from their networks [35-38]. The ability to measure how people engage and share misinformation on social media may help us better target and monitor the impact of public health interventions [39-41].

Given the rate at which new information is made available and the resources needed to appraise them, there is currently no way to keep up with new health-related stories as soon as they appear. Although the challenge of managing information volume versus quality was discussed two decades ago [42], methods for managing emerging misinformation in health-related news and media remain an unresolved issue for public health.

Research Objectives

We sought to characterize the sharing and potential reach of vaccine-related Web pages shared on Twitter, relative to credibility. As it would not have been feasible to manually assess the credibility of all Web pages, we developed and evaluated classifiers to automatically estimate their credibility.

Methods

Overview

The study used a retrospective observational design. To estimate the credibility of vaccine-related Web pages shared on Twitter, we collected text from vaccination-related Web pages by monitoring links from tweets that mentioned relevant keywords. We manually appraised the credibility of a sample of Web pages by applying a checklist-based appraisal tool, using the sample to train classifiers to predict a credibility score in unseen Web pages. Applying an ensemble classifier to the full set of Web pages collected as part of the surveillance, we examined patterns of sharing relative to credibility scores.

Datasets

We collected 6,591,566 English language, vaccine-related tweets and retweets from 1,860,662 unique Twitter users between January 17, 2017, and March 14, 2018, using the Twitter Search Application Programming Interface, using a set of predefined search terms (including “vaccin*,” “immunis*,” “vax*,” and “antivax*”). For all unique users posting vaccine-related tweets during the study period, we collected the lists of their followers to construct the social network.

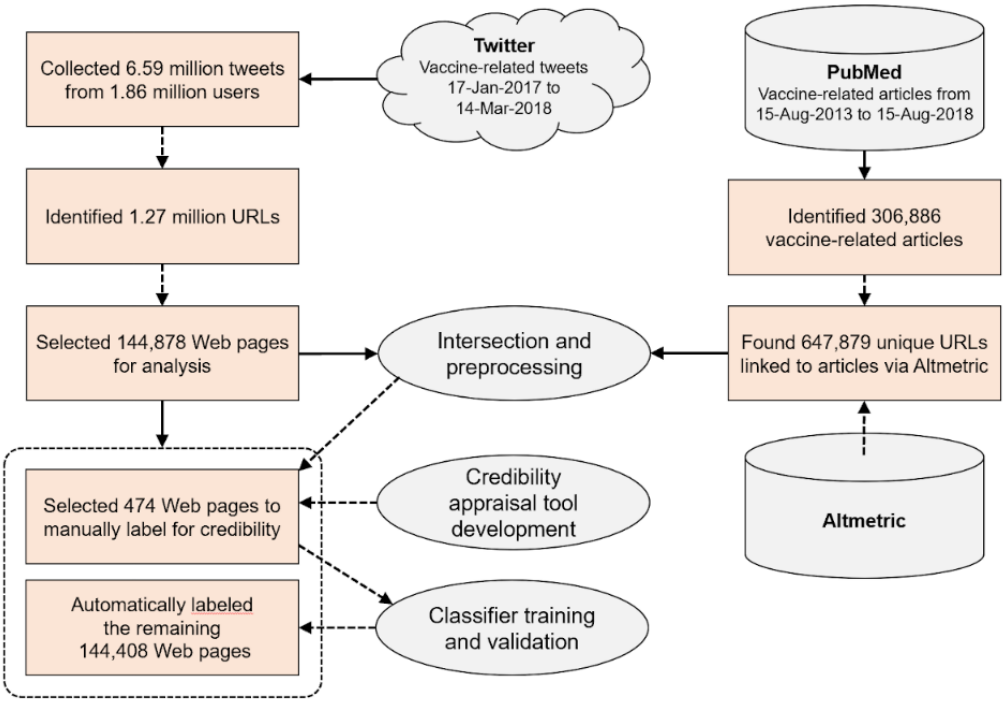
We extracted 1.27 million unique URLs from the set of tweets to identify the set of text-based Web pages to include in the analysis. To restrict the set of Web pages to only English language text, we used a Google library [43]; removed other Web pages that were internal Twitter links, broken links, or

links to Web pages that were no longer available; and removed Web pages with fewer than 300 words in contiguous blocks. We then checked for duplicates of other Web pages already included, removing Web pages for which most of the text was equivalent to another Web page in the set, retaining the Web page with the greatest number of words. The remaining set of 143,003 Web pages (Figure 1) was used in the subsequent analysis.

To modify how we sampled tweets for constructing a manually labeled dataset, we used PubMed to search for vaccine-related research articles using search terms "vaccine" or "immunisation" in the title or abstract, automatically expanded by PubMed to include synonyms and MeSH terms. The search returned 306,886 articles. We then used the PubMed identifiers of these articles with Altmetric (Digital Science) to identify Web pages (news, blogs, and social media posts) that linked to these articles via their digital object identifier, PubMed entry, or journal Web page. We found 647,879 unique URLs from Altmetric that cited the selected vaccines-related PubMed articles.

The intersection of the URLs extracted from Altmetric and the URLs extracted from the tweets allowed us to oversample from the set of Web pages for which we expected to have higher-credibility scores (described below). This approach also allowed us to exclude most of the URLs shared on Twitter that linked directly to research articles by removing the tweets that were identified by Altmetric.

Figure 1. The steps used to define the training dataset and automatically label Web pages.



Credibility Appraisal Tool

The credibility appraisal tool was developed by 3 investigators (AGD, AD, and MS) with expertise in public health, public health informatics, science communication, and journalism. To develop a tool that would work specifically with vaccine-related Web pages, the investigators adapted and synthesized individual criteria from the following checklist-based tools and guidelines [19]:

- Centers for Disease Control and Prevention guide for creating health materials [44]
- The DISCERN tool [21]
- Health News Review criteria [23] that is informed by Moynihan et al [45] and the Statement of Principles of the Association of Health Care Journalists [46]
- Media Doctor review criteria [47]
- World Health Organization report on vaccination and trust [48]
- The QIMR [22].

Using these documents as a guide, we adapted from the DISCERN and QIMR checklists, and added 2 additional criteria that were specific to vaccine-related communications. The tool was pilot tested on 30 randomly selected Web pages and iteratively refined through discussion among the 3 investigators. The resulting credibility appraisal tool included the following 7 criteria: (1) information presented is based on objective, scientific research; (2) adequate detail about the level of evidence offered by the research is included; (3) uncertainties and limitations in the research in focus are described; (4) the information does not exaggerate, overstate, or misrepresent available evidence; (5) provides context for the research in focus; (6) uses clear, nontechnical language that is easy to understand; and (6) is transparent about sponsorship and funding.

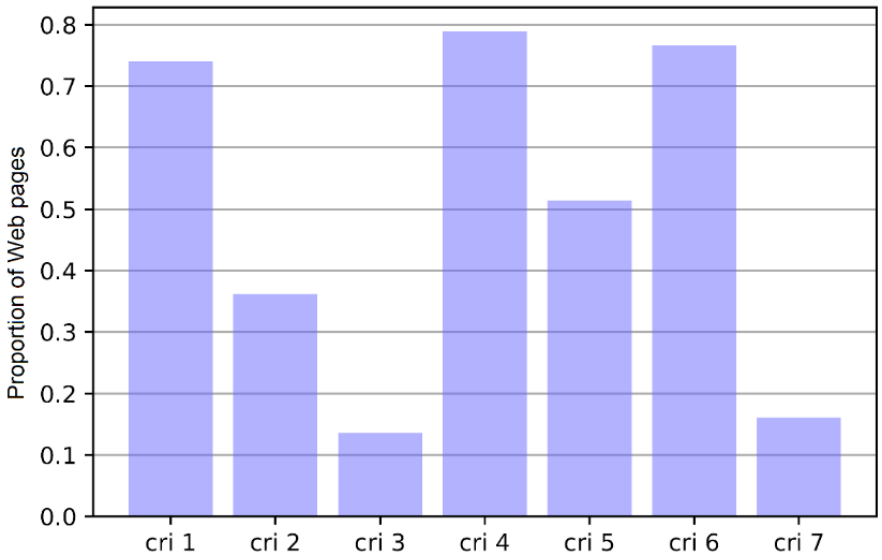
Manually Labeled Sample

The 3 investigators then applied the credibility appraisal tool to an additional 474 vaccine-related Web pages. For each Web page, investigators navigated to the website, read the article, and decided whether it satisfied each of the 7 criteria. This

process produced a set of values (0 or 1) for each criterion and Web page. We then summarized the information as a *credibility score*, defined by the number of criteria that were satisfied, and grouped Web pages by credibility score into low (0-2 criteria satisfied), medium (3-4 criteria satisfied), and high (5-7 criteria satisfied). Across the 474 expert-labeled examples, the proportion of the Web pages that were judged to have satisfied each of the 7 credibility criteria varied substantially (Figure 2).

The investigators independently undertook duplicate appraisals of a subset of the Web pages to measure inter-rater reliability, and it was found to be reasonable for separating Web pages as low, medium, or high credibility (Fleiss kappa 0.46; 95% CI 0.41-0.52; P<.001) and near-perfect when the aim was to separate low-credibility Web pages from all others (Fleiss kappa 0.89; 95% CI 0.82-0.97; P<.001). The design of the checklist suggests that it is a useful approach for identifying Web pages of low credibility.

Figure 2. The proportion of Web pages that met the individual criteria in the 474 Web pages used to train the classifiers. cri: criterion.



Classifier Design

We compared 3 machine learning methods that are commonly used for document classification problems: support vector machines (SVM), random forests (RF), and recurrent neural networks (RNN). The SVM method trains a large-margin classifier that aims to find a decision boundary between 2 classes that is maximally far from any point in the training data. In the RF method classification, trees are constructed by randomly selecting a subspace of features at each node of the decision tree to grow branches. The method then uses bagging to generate subsets of training data for constructing individual trees, which are then combined to form RF model. The RNN method refers to a class of artificial neural networks comprising neural network blocks that are linked to each other to form a directed graph along a sequence. The method is used to model dynamic temporal behavior for a time sequence, which is useful for understanding the language.

The aim of these supervised machine learning techniques was to train a model to predict the class of an unseen document by learning how to distinguish the language used across classes. To apply the classifiers, we cleaned the text downloaded from Web pages by removing extra spaces, tabs, extra newlines, and nonstandard characters including emoticons. Each Web page was then included as a document in our corpus.

To develop the RNN classifier, we used average-stochastic gradient descent weight-dropped long short-term memory [49]. In what follows, we refer to this as the deep learning (DL)–based classifier. The DL-based classifier comprised a backbone and custom head. The backbone is a language model that is a deep RNN. The head is a linear classifier comprising 2 linear blocks with rectified linear unit activations for the intermediate layer and a softmax activation for the final layer that can estimate the target labels (in our case, whether it satisfies a credibility criterion).

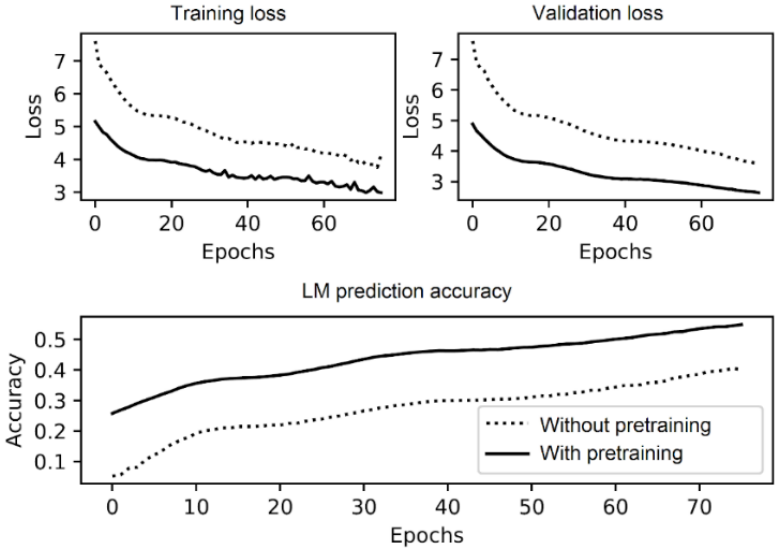
Language models are trained to understand the structure of the language used in a corpus of documents, and its performance is measured by its ability to predict the next word in a sentence based on the set of previous words. After the language model is trained for this task, the complete DL-based classifier is then fine-tuned to predict whether a document satisfies each of the credibility checklist criteria. Language models are often trained to learn the structure of the language in a target corpus, but recent advances in transfer learning have produced superior results including shorter training times and higher performance. An example is the Universal Language Model Fine-Tuning method [50], which was proposed and evaluated on natural language processing tasks.

We used transfer learning to create the language model backbone. The language model was developed with 3 layers, 1150 hidden units, and an embedding size of 400 per word, and the weights were initialized from a pretrained WikiText-103 language model produced by Howard et al [50]. The parameters and values used in the initialization of the language model and classifier are given in Table 1. The results of the performance of the associated language model are given in Figure 3.

Table 1. The parameters and corresponding values for the initialization of the language model and classifier.

Parameters	Value
Weight decay	1.00E-04
Backpropagation through time	60
Batch size	52
Dropouts	0.25, 0.1, 0.2, 0.02, 0.15
Embedding size	400
Number of layers	3 (language model), 5 (classifier)
Optimizer	Adam
β_1, β_2	0.8, 0.99

Figure 3. The performance difference of the language model (LM) for 2 different settings, including training loss (top-left), validation cross-entropy loss (top-right), and the accuracy of the LM predicting the next word in a sentence given previous words in the validation text (bottom).



For the SVM- and RF-based classifiers, we performed additional preprocessing to remove stop words and low-frequency words to improve accuracy. After preprocessing, there were 60,660 unique words used across the entire corpus; these were used as features for training and testing RF and SVM classifiers. Each document was represented as a set of feature vectors, where features were defined by term frequency–inverse document frequency (tf-idf) weights. tf-idf represents the importance of a word to a document in a corpus, which increases proportionally to the number of times it appears in the document but is offset by the frequency of the word in the corpus, ensuring that the similarity between documents be more influenced by discriminative words with relatively low frequencies in the corpus. The best parameters for SVM and RF are found using grid search functionality of *scikit-learn* library and are given in Table 2.

Using the expert-labeled data, we trained 21 classifiers (1 per criterion for each of the RF-, SVM-, and DL-based classifiers) and evaluated the performance of the classifiers in 10-fold cross-validation tests, reporting the average F_1 score and accuracy for all 3 classifiers. Although the comparison of the performance across the set of classifiers may be of interest, our aim was to provide the basis for an ensemble classifier that could reliably estimate which of the criteria were met by each Web page.

Table 2. The parameters used for support vector machine and random forest classifiers; all other parameters are kept as default.

Parameters	Value
Support vector machines	
C	100
Gamma	1
Kernel	linear
Norm	l1
Use-idf[a]	TRUE
Max-df[b]	1
N-gram range	(1,1)
Random forests	
N-estimators	10
Criterion	Gini
Min-impurity-split	1.00E-07

[a]Use-idf: when true, term weights are scaled by the number of documents they appear in.

[b]Max-df: when set to 1, words that appear in every document are not removed.

Sharing and Potential Exposure Estimation

Following the development of a reliable tool for automatically estimating the credibility of vaccine-related communications at scale, we aimed to characterize patterns of potential exposure to low-credibility vaccine communications on Twitter. For each Web page that met our study inclusion criteria, we estimated its credibility score using the best-performing classifiers for each criterion. We then aggregated the total number of tweets posted during the study period that included a link to the Web page, including tweets and retweets. We then estimated the *potential exposure* by summing the total number of followers for all tweets and retweets. Note that this represents the maximum possible audience, and we did not identify the unique set of users who might have been exposed at least once because of who they follow as we had done in previous studies [14].

To examine how users posting links to low-credibility Web pages might be concentrated within or across subpopulations, we also estimated a per-user measure of credibility, which was defined by the list of credibility scores for any user sharing links to one or more Web pages. We used these lists in conjunction with information about followers to construct a *follower network*, which allowed us to identify subpopulations of Twitter users for which the sharing of low-credibility vaccine communications was common.

Results

Classifier Performance

The RF classifiers produced the highest performance overall, and in most cases predicted, whether the text on a vaccine-related Web page satisfied each of the credibility criteria with over 90% accuracy (Table 3). The SVM-based classifier produced the highest F_1 scores for 2 of the most unbalanced criteria. Further experiments are needed to determine whether the DL-based classifier outperforms baseline methods if more expert-labeled data are made available. The results show that it is feasible to estimate credibility appraisal for Web pages about vaccination without additional human input, suggesting the performance—although variable—is high enough to warrant their use in surveillance.

Where the best-performing classifiers were combined to distinguish between low-, medium-, and high-credibility Web pages, the overall accuracy of the ensemble classifier that combines best-performing classifiers (SVM for criterion 3 and 7 and RF for all other criteria) was 78.30%. In terms of labeling low-credibility Web pages, the ensemble classifier rarely mislabeled a high- or medium-credibility Web page as low credibility; more than 19 out of every 20 Web pages labeled as low credibility were correct.

To consider the expected robustness of the classifiers, we additionally analyzed the set of terms that were most informative of low-credibility Web pages. We used a Fisher exact test to compare the proportion of low-credibility Web pages a term appeared in at least once relative to the proportion of other Web pages in which the term appeared at least once, examining the terms that were over-represented in either direction (Figure 4).

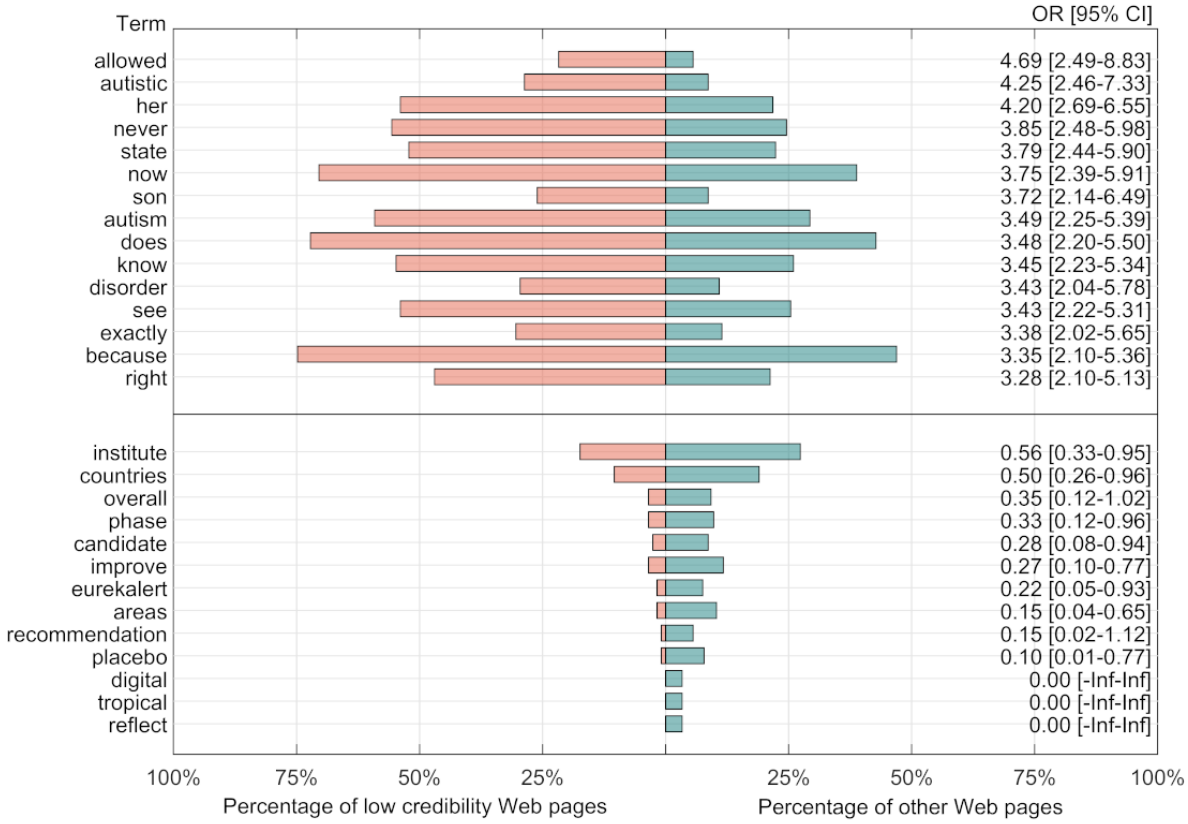
The results indicate a set of mostly general terms; terms that are most indicative of low-credibility Web pages are related to stories about individuals and individual autonomy (eg, “her,” “son,” “autistic,” “right,” and “allowed”), and terms that are most indicative of other Web pages are related to research and populations (eg, “institute,” “phase,” “placebo,” “countries,” “improve,” and “tropical”). The results suggest that the sample of Web pages used to construct the training data is a broad enough sample to capture general patterns rather than specific repeated topics that would limit the external validity of the approach.

Table 3. Performance of the classifiers (average F_1 score and accuracy in 10-fold cross-validation).

Criterion	Deep learning[a], mean (SD)		Support vector machines[a], mean (SD)		Random forests[a], mean (SD)	
	F_1 score	Accuracy	F_1 score	Accuracy	F_1 score	Accuracy
1	0.851 (0.005)	0.740 (0.008)	0.903 (0.032)	0.842 (0.045)	*0.950* (0.015)	0.924 (0.019)
2	0.000 (0.000)	0.638 (0.003)	0.802 (0.044)	0.828 (0.018)	*0.915* (0.005)	0.943 (0.006)
3	0.000 (0.000)	0.865 (0.009)	*0.761* (0.038)	0.917 (0.011)	0.745 (0.088)	0.944 (0.018)
4	0.882 (0.001)	0.789 (0.002)	0.903 (0.042)	0.833 (0.068)	*0.959* (0.017)	0.936 (0.022)
5	0.551 (0.249)	0.486 (0.051)	0.787 (0.034)	0.721 (0.051)	*0.921* (0.022)	0.920 (0.020)
6	0.867 (0.002)	0.765 (0.004)	0.912 (0.006)	0.852 (0.010)	*0.964* (0.002)	0.943 (0.004)
7	0.000 (0.000)	0.840 (0.008)	*0.801* (0.029)	0.924 (0.006)	0.764 (0.057)	0.936 (0.004)

[a]The classifier with the highest F1-score is italicized for each criterion.

Figure 4. A subset of the terms that were informative of low-credibility scores in the training set of 474 Web pages. Terms at the top are those most over-represented in low-credibility Web pages compared with other Web pages, and terms at the bottom are those most under-represented in low-credibility Web pages compared with other Web pages. OR: odds ratio; Inf: infinity.



Potential Exposure Estimation

Satisfied with the performance of the ensemble classifier, we then applied it to the full set of 144,003 unique vaccine-related Web pages, producing an estimated credibility score for every page. Fewer Web pages with low-credibility scores were shared on Twitter relative to those with medium- or high-credibility scores (Figure 5), although it is important to consider the performance limitations of the ensemble classifier when interpreting these findings. We estimated that 11.86% (16,961 of 143,003) of Web pages were of low credibility, and they generated 14.68% (112,225 of 764,283) of retweets. In comparison, 23.52% (33,636 of 143,003) of Web pages were of high credibility, and they generated 21.04% (160,777 of 764,283) of all retweets.

When we examined the total number of potential exposures by counting cumulative followers across all tweets and retweets for each Web page, we found that the distributions were similar (illustrated by the slopes of the 3 distributions in Figure 6).

Figure 5. The sum of tweets and retweets for links to included Web pages relative to the number of credibility criteria satisfied.

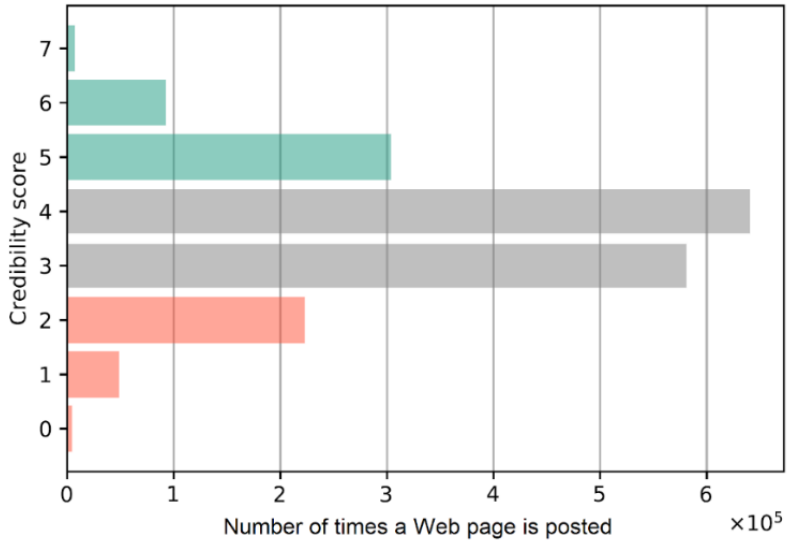
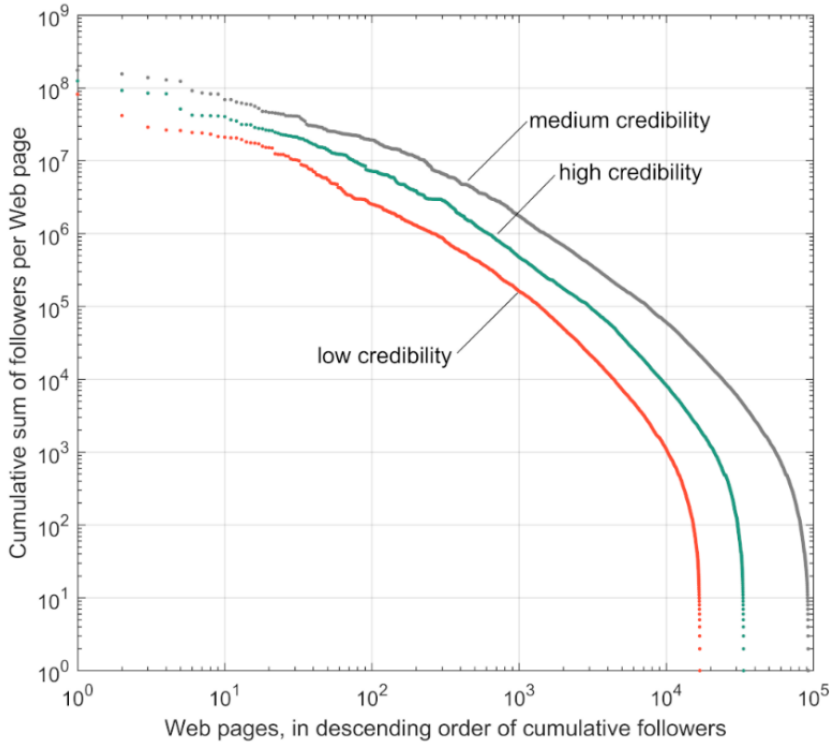


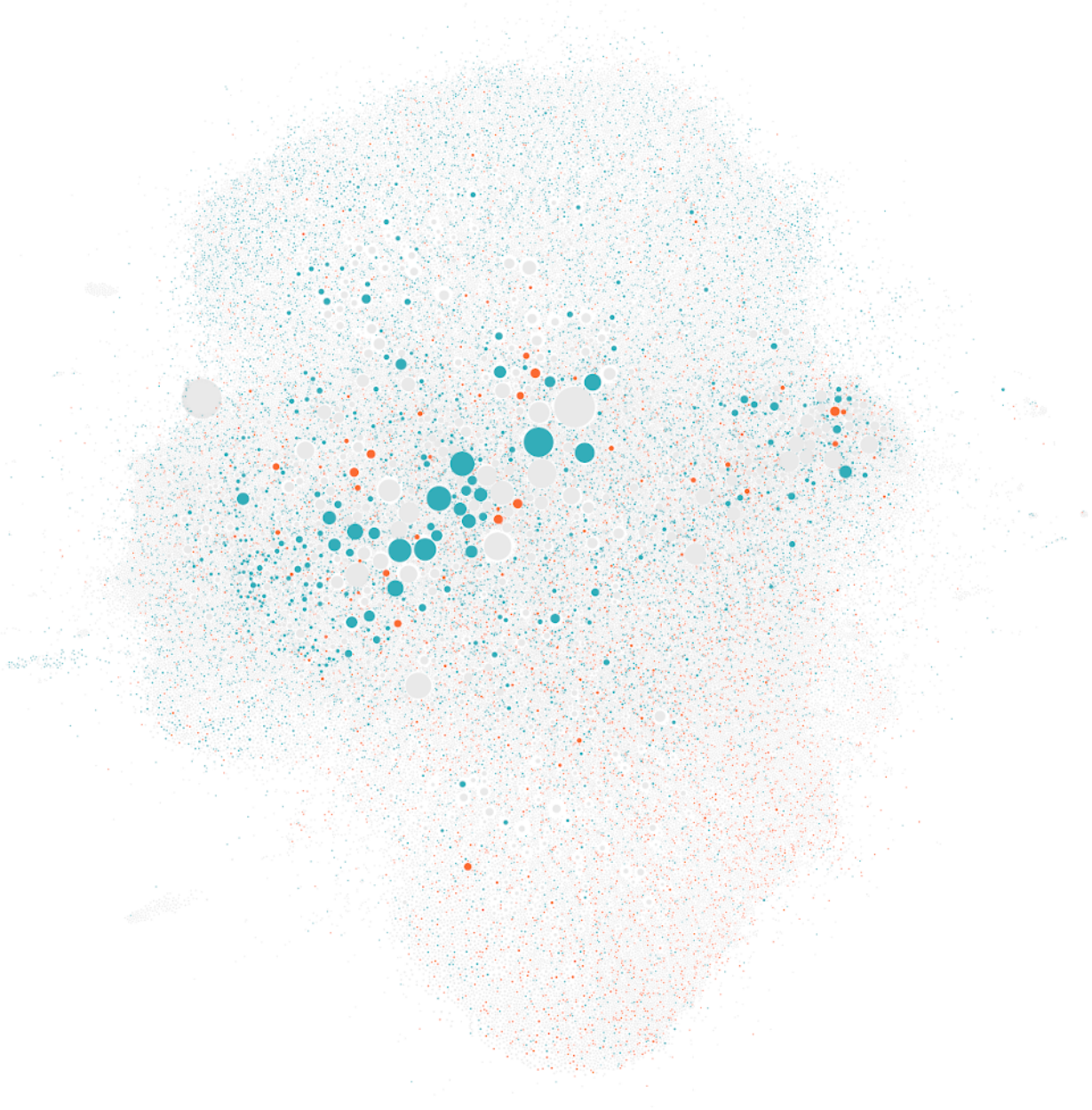
Figure 6. The distribution of potential exposures per Web page for low (orange), medium (gray), and high (cyan) credibility scores, where low credibility includes scores from 0 to 2, and high credibility includes scores from 5 to 7.



Measured by the total proportion of exposures to links to relevant Web pages, tweets to low credibility Web pages produced 9.34% (1.64 billion of 17.6 billion) of total exposures, compared with the 24.59% (4.33 billion of 17.6 billion) of total exposures to high-credibility Web pages. This indicates that Twitter users sharing links to high-credibility and medium-credibility vaccine-related Web pages tended to have a greater number of followers than those sharing links to low-credibility vaccine-related Web pages. However, the shape of the distribution shows that some of the low-credibility Web pages may have been influential; the top 100 Web pages by exposure were included in tweets that may have been seen by 2 million to 80 million users, and more than 200 Web pages of low credibility were included in tweets that could have reached 1 million users.

Links to low-credibility vaccine-related Web pages were more heavily concentrated among certain groups of users posting tweets about vaccines on Twitter. This is evident in a visualization of the follower network for the set of 98,663 Twitter users who posted at least two links to Web pages included in the study (Figure 7). The network indicates heterogeneity in the sharing of links to low-credibility vaccine-related Web pages, suggesting that there are likely to be communities of social media users for whom the majority of what they see and read about vaccines is of low credibility.

Figure 7. A network visualization representing the subset of 98,663 Twitter users who posted tweets including links to vaccine-related Web pages at least twice and were connected to at least one other user in the largest connected component. Users who posted at least 2 high-credibility Web pages and no low-credibility Web pages (cyan) and those who posted at least two low-credibility Web pages and no high-credibility Web pages (orange) are highlighted. The size of the nodes is proportional to the number of followers each user has on Twitter, and nodes are positioned by a heuristic such that well-connected groups of users are more likely to be positioned close together in the network diagram.



Discussion

Principal Findings

We found that it is feasible to produce machine learning classifiers to identify vaccine-related Web pages of low credibility. Applying a classifier to vaccine-related Web pages shared on Twitter between January 2017 and March 2018, we found that fewer low-credibility Web pages were shared overall, though some had a potential reach of tens of millions of Twitter users. A network visualization suggested that certain communities of Twitter users were much more likely to share and be exposed to low-credibility Web pages.

Research in Context

This research extends knowledge related to the surveillance of health misinformation on social media. Where much of the prior research has aimed to label individual social media posts or the claims made on social media by veracity [25-29], we instead labeled Web pages shared on social media using a credibility appraisal checklist extended from previously validated instruments to be appropriate to vaccine-related communications [21,22]. In other related work, Mitra et al [51] examined the linguistic features in social media posts that influenced perceptions of credibility. Although we did not examine the linguistic features of the tweets that included links to low-credibility information, it would be interesting to connect these ideas to better understand whether they influence user

behavior—making users more likely to engage with a tweet by URL access, replying, and sharing.

The work presented here is also different from previous studies examining opinions and attitudes expressed by Twitter users, which mostly label individual tweets or users based on whether they are promoting vaccination or advocating against vaccines [30,32,35,38]. Here we consider the communications shared on Twitter rather than the opinions expressed by users in the text of tweets.

Our study is also not directly comparable with previous studies that have examined how misinformation spreads through social media [2-6]. We examined a single topic that might not generalize to other application domains such as politics, labeled information according to a broader set of criteria than just the veracity of the information, and measured total potential exposures rather than just cascades of tweets and retweets. Rather than sampling from a set of known examples of fake and real news to compare spread, we sampled from across the spectrum of relevant articles shared on Twitter. Structuring the experiments in this way, we found no clear difference in the distribution of total potential exposures between low-credibility Web pages and others. Although most low-credibility Web pages are shared with a smaller number of Twitter users, some had the potential to reach tens of millions.

Implications

This study has implications for public health. The ability to measure how people engage with and share misinformation on social media may help us better target and monitor the impact of public health interventions [39-41]. We found that certain subpopulations of Twitter users share low-credibility vaccine communications more often and are less likely to be connected to users sharing higher-credibility vaccine communications. Although these results are unsurprising, most studies examining vaccines on social media have only counted tweets rather than examining the heterogeneity of potential exposure to vaccine critical posts [30,38,52], despite evidence of the clustering of opinions from as early as 2011 [32]. This study is consistent with these previous findings on clustering and studies examining exposure to different topics about HPV vaccines [35,37]. Knowing where low-credibility communications are most commonly shared on social media may support the development of communication interventions targeted specifically at communities that are most likely to benefit [53]. Although the methods are not yet precise enough to reliably identify individual links to low-credibility communications, they may eventually be useful as the basis for countermeasures such as active debunking. Methods for inoculating against misinformation by providing warnings immediately before access have mixed results [10,54,55].

Limitations

There were several limitations to this study. Although we used a modified sampling strategy to ensure a more balanced representation of Web pages, the manually labeled sample used for training and internal validation was relatively small, and this might have affected the results in 2 ways. First, our results showed that the DL-based classifiers were less accurate than the RF-based classifiers, but this might have been the consequence of the available training data rather than the general value of the DL approach. Without testing on larger sets of training data, we are unable to reliably conclude about the comparative performance of the machine learning methods. Second, in some document classification tasks where features are relatively sparse or many documents are very similar, using a smaller set of labeled examples can lead to overfitting. To avoid this, we were careful about removing duplicates and Web pages with overlapping text.

A second type of limitation relates to the choices we made about the methods. Other methods and architectures could have been used to predict credibility from text. For example, we could have used simpler methods including Naïve Bayes and logistic regression, used a single multi-label classifier to predict whether a document extracted from a Web page satisfied any of the criteria, or constructed a model that directly predicts the credibility score rather than the individual components.

A further limitation relates to the external validity of the classifier and our inability to draw conclusions about Web pages that do not include contiguous sections of text. We included only Web pages from which we could extract contiguous blocks of text and used a novel approach to sampling from those Web pages to create a reasonably balanced sample across the set of credibility scores. Other URLs included in vaccine-related tweets included links to other social media posts (including links to other tweets), links to YouTube and Instagram, links to memes in which text is embedded in an image, links to dynamic pages that no longer show the same information, and links to a range of other pages that included videos or images alongside a small amount of text. As we were unable to estimate the credibility of the vaccine-related information presented on these other Web pages, our conclusions are limited to the characterization of text-based Web pages. It is likely that a substantial proportion of Instagram, Facebook, and YouTube Web pages would receive a low-credibility score if they were assessed [56-58], which means we may have underestimated the sharing of low-credibility vaccine-related communications on Twitter.

Our estimates of exposure were imperfect. To estimate how many Twitter users might have been exposed to information relative to credibility, we summed the total number of followers of a user for each user that posted the link. We did not count the total number of unique followers who might have seen the link, did not report the number of likes, and do not have access to the number of replies. In the absence of more detailed measures of engagement that can estimate the number of times a Web page was accessed via Twitter, we felt measures of potential exposure were a reasonable upper bound. The conclusions related to measures of potential exposure, therefore, need to be interpreted with caution, and further studies using robust epidemiological designs are needed to reliably estimate exposure.

Conclusions

We developed and tested machine learning methods to support the automatic credibility appraisal of vaccine-related information on the Web, showing that it is feasible. This allowed us to scale

our analysis of large-scale patterns of potential exposure to low-credibility vaccine-related Web pages shared on Twitter. We found that although low-credibility Web pages were shared less often overall, there were certain subpopulations where the sharing of low-credibility Web pages was common. The results suggest two new ways to address the challenge of misinformation, including ongoing surveillance to identify at-risk communities and better target resources in health promotion and embedding the tool in interventions that flag low-credibility communications for consumers as they engage with links to Web pages on social media.

Acknowledgments

The study was funded by the National Health & Medical Research Council Project Grant APP1128968. The authors would like to thank Paige Martin for contributing to the research with the management of the database, data collection, and data synthesis and Maryke Steffens for the design and development of the credibility criteria and expert labeling.

Conflicts of Interest

None declared.

References

1. Lazer DM, Baum MA, Benkler Y, Berinsky AJ, Greenhill KM, Menczer F, et al. The science of fake news. Science 2018 Mar 9;359(6380):1094-1096. [doi: 10.1126/science.aao2998] [Medline: 29590025]
2. Budak C, Agrawal D, El AA. Limiting the Spread of Misinformation in Social Networks. In: Proceedings of the 20th International Conference on World Wide Web. 2011 Presented at: WWW'11; March 28-April 1, 2011; Hyderabad, India p. 665-674. [doi: 10.1145/1963405.1963499]
3. Mocanu D, Rossi L, Zhang Q, Karsai M, Quattrociocchi W. Collective attention in the age of (mis)information. Comput Hum Behav 2015 Oct;51:1198-1204. [doi: 10.1016/j.chb.2015.01.024]
4. Tambuscio M, Ruffo G, Flammini A, Menczer F. Fact-Checking Effect on Viral Hoaxes: A Model of Misinformation Spread in Social Networks. In: Proceedings of the 24th International Conference on World Wide Web. 2015 Presented at: WWW'15; May 18-22, 2015; Florence, Italy p. 977-982. [doi: 10.1145/2740908.2742572]
5. Kumar S, West R, Leskovec J. Disinformation on the Web: Impact, Characteristics, and Detection of Wikipedia Hoaxes. In: Proceedings of the 25th International Conference on World Wide Web. 2016 Presented at: WWW '16; April 11-15, 2016; Montreal, Canada p. 591-602. [doi: 10.1145/2872427.2883085]
6. Vosoughi S, Roy D, Aral S. The spread of true and false news online. Science 2018 Mar 9;359(6380):1146-1151. [doi: 10.1126/science.aap9559] [Medline: 29590045]
7. Grinberg N, Joseph K, Friedland L, Swire-Thompson B, Lazer D. Fake news on Twitter during the 2016 US presidential election. Science 2019 Jan 25;363(6425):374-378. [doi: 10.1126/science.aau2706] [Medline: 30679368]
8. del Vicario M, Bessi A, Zollo F, Petroni F, Scala A, Caldarelli G, et al. The spreading of misinformation online. Proc Natl Acad Sci U S A 2016 Jan 19;113(3):554-559 [FREE Full text] [doi: 10.1073/pnas.1517441113] [Medline: 26729863]
9. Weaver JB, Thompson NJ, Weaver SS, Hopkins GL. Healthcare non-adherence decisions and internet health information. Comput Hum Behav 2009 Nov;25(6):1373-1380. [doi: 10.1016/j.chb.2009.05.011]
10. Lewandowsky S, Ecker UK, Seifert CM, Schwarz N, Cook J. Misinformation and its correction: continued influence and successful debiasing. Psychol Sci Public Interest 2012 Dec;13(3):106-131. [doi: 10.1177/1529100612451018] [Medline: 26173286]
11. Schwitzer G. A review of features in internet consumer health decision-support tools. J Med Internet Res 2002;4(2):E11 [FREE Full text] [doi: 10.2196/jmir.4.2.e11] [Medline: 12554558]
12. Eysenbach G, Köhler C. How do consumers search for and appraise health information on the world wide web? Qualitative study using focus groups, usability tests, and in-depth interviews. Br Med J 2002 Mar 9;324(7337):573-577 [FREE Full text] [doi: 10.1136/bmj.324.7337.573] [Medline: 11884321]
13. Reuters Institute Digital News Report. 2019. Reuters Institute for the Study of Journalism, University of Oxford URL: http://www.digitalnewsreport.org/ [accessed 2019-06-29]
14. Winters M, Larsson A, Kowalski J, Sundberg CJ. The association between quality measures of medical university press releases and their corresponding news stories-important information missing. PLoS One 2019;14(6):e0217295 [FREE Full text] [doi: 10.1371/journal.pone.0217295] [Medline: 31188838]
15. Yavchitz A, Boutron I, Bafeta A, Marroun I, Charles P, Mantz J, et al. Misrepresentation of randomized controlled trials in press releases and news coverage: a cohort study. PLoS Med 2012;9(9):e1001308 [FREE Full text] [doi: 10.1371/journal.pmed.1001308] [Medline: 22984354]
16. Haneef R, Ravaud P, Baron G, Ghosn L, Boutron I. Factors associated with online media attention to research: a cohort study of articles evaluating cancer treatments. Res Integr Peer Rev 2017;2:9 [FREE Full text] [doi: 10.1186/s41073-017-0033-z] [Medline: 29451556]

17. Sumner P, Vivian-Griffiths S, Boivin J, Williams A, Venetis CA, Davies A, et al. The association between exaggeration in health related science news and academic press releases: retrospective observational study. Br Med J 2014 Dec 9;349:g7015 [FREE Full text] [doi: 10.1136/bmj.g7015] [Medline: 25498121]
18. Grundy Q, Dunn AG, Bourgeois FT, Coiera E, Bero L. Prevalence of disclosed conflicts of interest in biomedical research and associations with journal impact factors and altmetric scores. J Am Med Assoc 2018 Jan 23;319(4):408-409 [FREE Full text] [doi: 10.1001/jama.2017.20738] [Medline: 29362787]
19. Bernstam EV, Shelton DM, Walji M, Meric-Bernstam F. Instruments to assess the quality of health information on the World Wide Web: what can our patients actually use? Int J Med Inform 2005 Jan;74(1):13-19. [doi: 10.1016/j.ijmedinf.2004.10.001] [Medline: 15626632]
20. Steffens M, Dunn A, Leask J. Meeting the challenges of reporting on public health in the new media landscape. Aust J Rev 2017;39(2):119-132 [FREE Full text]
21. Charnock D, Shepperd S, Needham G, Gann R. DISCERN: an instrument for judging the quality of written consumer health information on treatment choices. J Epidemiol Community Health 1999 Feb;53(2):105-111 [FREE Full text] [doi: 10.1136/jech.53.2.105] [Medline: 10396471]
22. Zeraatkar D, Obeda M, Ginsberg JS, Hirsh J. The development and validation of an instrument to measure the quality of health research reports in the lay media. BMC Public Health 2017 Apr 20;17(1):343 [FREE Full text] [doi: 10.1186/s12889-017-4259-y] [Medline: 28427426]
23. HealthNewsReview. Our Review Criteria URL: https://www.healthnewsreview.org/about-us/review-criteria/ [accessed 2019-06-29]
24. Larson H. The biggest pandemic risk? Viral misinformation. Nature 2018 Oct;562(7727):309. [doi: 10.1038/d41586-018-07034-4] [Medline: 30327527]
25. Qazvinian V, Rosengren E, Radev D, Mei Q. Rumor Has It: Identifying Misinformation in Microblogs. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing. 2011 Presented at: EMNL '11; July 27-31, 2011; Edinburgh, United Kingdom p. 1589-1599.
26. Zhao Z, Resnick P, Mei Q. Enquiring Minds: Early Detection of Rumors in Social Media from Enquiry Posts. In: Proceedings of the 24th International Conference on World Wide Web. 2015 Presented at: WWW'15; May 18-22, 2015; Florence, Italy p. 1395-1405. [doi: 10.1145/2736277.2741637]
27. Vosoughi S, Mohsenvand M, Roy D. Rumor gauge: predicting the veracity of rumors on Twitter. ACM Trans Knowl Discov Data 2017 Aug 21;11(4):1-36. [doi: 10.1145/3070644]
28. Ma J, Gao W, Mitra P, Kwon S, Jansen BJ, Wong KF, et al. Detecting Rumors From Microblogs With Recurrent Neural Networks. In: Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence. 2016 Presented at: IJCAI'16; July 09-15, 2016; New York, United States p. 3818-3824.
29. Ruchansky N, Seo S, Liu Y. CSI: A Hybrid Deep Model for Fake News Detection. In: Proceedings of the 2017 ACM on Conference on Information and Knowledge Management. 2017 Presented at: CIKM'17; November 06-10, 2017; Singapore, Singapore p. 797-806. [doi: 10.1145/3132847.3132877]
30. Broniatowski DA, Jamison AM, Qi S, AlKulaib L, Chen T, Benton A, et al. Weaponized health communication: Twitter bots and Russian trolls amplify the vaccine debate. Am J Public Health 2018 Oct;108(10):1378-1384. [doi: 10.2105/AJPH.2018.304567] [Medline: 30138075]
31. Dredze M, Broniatowski DA, Smith MC, Hilyard KM. Understanding vaccine refusal: why we need social media now. Am J Prev Med 2016 Apr;50(4):550-552 [FREE Full text] [doi: 10.1016/j.amepre.2015.10.002] [Medline: 26655067]
32. Salathé M, Khandelwal S. Assessing vaccination sentiments with online social media: implications for infectious disease dynamics and control. PLoS Comput Biol 2011 Oct;7(10):e1002199 [FREE Full text] [doi: 10.1371/journal.pcbi.1002199] [Medline: 22022249]
33. Du J, Tang L, Xiang Y, Zhi D, Xu J, Song H, et al. Public perception analysis of tweets during the 2015 measles outbreak: comparative study using convolutional neural network models. J Med Internet Res 2018 Jul 9;20(7):e236 [FREE Full text] [doi: 10.2196/jmir.9413] [Medline: 29986843]
34. Chew C, Eysenbach G. Pandemics in the age of Twitter: content analysis of tweets during the 2009 H1N1 outbreak. PLoS One 2010 Nov 29;5(11):e14118 [FREE Full text] [doi: 10.1371/journal.pone.0014118] [Medline: 21124761]
35. Dunn AG, Leask J, Zhou X, Mandl KD, Coiera E. Associations between exposure to and expression of negative opinions about human papillomavirus vaccines on social media: an observational study. J Med Internet Res 2015 Jun 10;17(6):e144 [FREE Full text] [doi: 10.2196/jmir.4343] [Medline: 26063290]
36. Dunn AG, Surian D, Leask J, Dey A, Mandl KD, Coiera E. Mapping information exposure on social media to explain differences in HPV vaccine coverage in the United States. Vaccine 2017 May 25;35(23):3033-3040 [FREE Full text] [doi: 10.1016/j.vaccine.2017.04.060] [Medline: 28461067]
37. Surian D, Nguyen DQ, Kennedy G, Johnson M, Coiera E, Dunn AG. Characterizing Twitter discussions about HPV vaccines using topic modeling and community detection. J Med Internet Res 2016 Aug 29;18(8):e232 [FREE Full text] [doi: 10.2196/jmir.6045] [Medline: 27573910]

38. Du J, Cunningham RM, Xiang Y, Li F, Jia Y, Boom JA, et al. Leveraging deep learning to understand health beliefs about the Human Papillomavirus Vaccine from social media. NPJ Digit Med 2019;2:27 [FREE Full text] [doi: 10.1038/s41746-019-0102-4] [Medline: 31304374]
39. Shao C, Ciampaglia G, Flammini A, Menczer F. Hoaxy: A Platform for Tracking Online Misinformation. In: Proceedings of the 25th International Conference Companion on World Wide Web. 2016 Presented at: WWW'16; April 11-15, 2016; Montréal, Canada p. 745-750. [doi: 10.1145/2872518.2890098]
40. Dunn AG, Mandl KD, Coiera E. Social media interventions for precision public health: promises and risks. NPJ Digit Med 2018;1:47 [FREE Full text] [doi: 10.1038/s41746-018-0054-0] [Medline: 30854472]
41. Colditz JB, Chu K, Emery SL, Larkin CR, James AE, Welling J, et al. Toward real-time infoveillance of Twitter health messages. Am J Public Health 2018 Aug;108(8):1009-1014. [doi: 10.2105/AJPH.2018.304497] [Medline: 29927648]
42. Coiera E. Information epidemics, economics, and immunity on the internet. We still know so little about the effect of information on public health. Br Med J 1998 Nov 28;317(7171):1469-1470 [FREE Full text] [doi: 10.1136/bmj.317.7171.1469] [Medline: 9831570]
43. Google Code. 2018. Language Detection URL: https://code.google.com/p/language-detection/ [accessed 2019-06-29]
44. Centers for Disease Control and Prevention. 2009. Simply Put; A Guide for Creating Easy-to-understand Materials. Third Edition URL: https://stacks.cdc.gov/view/cdc/11938 [accessed 2019-06-29]
45. Moynihan R, Bero L, Ross-Degnan D, Henry D, Lee K, Watkins J, et al. Coverage by the news media of the benefits and risks of medications. N Engl J Med 2000 Jun 1;342(22):1645-1650. [doi: 10.1056/NEJM200006013422206] [Medline: 10833211]
46. Association of Health Care Journalists. Statement of Principles of the Association of Health Care Journalists URL: https://healthjournalism.org/secondarypage-details.php?id=56 [accessed 2019-06-29]
47. Wiggers J. The University of Newcastle. Media Doctor URL: https://www.newcastle.edu.au/research-and-innovation/centre/chb/research/media-doctor [accessed 2019-06-29]
48. Betsch C, Rossmann C, Habersaat KB, Pfeiffer D, Holtmann C, Korn L. World Health Organization Europe. 2017. Vaccination and Trust (2017) URL: http://www.euro.who.int/en/health-topics/disease-prevention/vaccines-and-immunization/publications/2017/vaccination-and-trust-2017 [accessed 2019-06-29]
49. Merity S, Keskar NS, Socher R. arXiv. 2017. Regularizing and optimizing LSTM language models URL: https://arxiv.org/abs/1708.02182 [accessed 2019-09-19]
50. Howard J, Ruder S. Universal Language Model Fine-Tuning for Text Classification. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. 2018 Presented at: ACL'18; 2018; Melbourne, Australia p. 328-339. [doi: 10.18653/v1/P18-1031]
51. Mitra T, Wright G, Gilbert E. A Parsimonious Language Model of Social Media Credibility Across Disparate Events. In: Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing. 2017 Presented at: CSCW'17; February 25-March 1, 2017; Portland, United States p. 126-145. [doi: 10.1145/2998181.2998351]
52. Tomeny TS, Vargo CJ, El-Toukhy S. Geographic and demographic correlates of autism-related anti-vaccine beliefs on Twitter, 2009-15. Soc Sci Med 2017 Oct;191:168-175 [FREE Full text] [doi: 10.1016/j.socscimed.2017.08.041] [Medline: 28926775]
53. Vraga EK, Bode L. Using expert sources to correct health misinformation in social media. J Sci Commun 2017 Sep 14;39(5):621-645. [doi: 10.1177/1075547017731776]
54. Bode L, Vraga EK. See something, say something: correction of global health misinformation on social media. Health Commun 2018 Sep;33(9):1131-1140. [doi: 10.1080/10410236.2017.1331312] [Medline: 28622038]
55. Pennycook G, Cannon TD, Rand DG. Prior exposure increases perceived accuracy of fake news. J Exp Psychol Gen 2018 Dec;147(12):1865-1880. [doi: 10.1037/xge0000465] [Medline: 30247057]
56. Briones R, Nan X, Madden K, Waks L. When vaccines go viral: an analysis of HPV vaccine coverage on YouTube. Health Commun 2012;27(5):478-485. [doi: 10.1080/10410236.2011.610258] [Medline: 22029723]
57. Madathil KC, Rivera-Rodriguez AJ, Greenstein JS, Gramopadhye AK. Healthcare information on YouTube: a systematic review. Health Informatics J 2015 Sep;21(3):173-194. [doi: 10.1177/1460458213512220] [Medline: 24670899]
58. Venkatraman A, Garg N, Kumar N. Greater freedom of speech on web 2.0 correlates with dominance of views linking vaccines to autism. Vaccine 2015 Mar 17;33(12):1422-1425. [doi: 10.1016/j.vaccine.2015.01.078] [Medline: 25665960]

Abbreviations

DL: deep learning
HPV: human papillomavirus
QIMR: Quality Index for health-related Media Reports
RF: random forests
RNN: recurrent neural networks
SVM: support vector machines
tf-idf: term frequency–inverse document frequency

Edited by G Eysenbach; submitted 17.03.19; peer-reviewed by MY Song, A Majmundar, A Louren, C Liang; comments to author 06.05.19; revised version received 29.06.19; accepted 02.09.19; published 04.11.19.

Please cite as:
Shah Z, Surian D, Dyda A, Coiera E, Mandl KD, Dunn AG
Automatically Appraising the Credibility of Vaccine-Related Web Pages Shared on Social Media: A Twitter Surveillance Study
J Med Internet Res 2019;21(11):e14007
URL: https://www.jmir.org/2019/11/e14007
doi:10.2196/14007
PMID:31682571

Original Paper

A Revised Model of Trust in Internet-Based Health Information and Advice: Cross-Sectional Questionnaire Study

Elizabeth Sillence[1], PhD; John Matthew Blythe[1,2], PhD; Pam Briggs[1], PhD; Mark Moss[3], PhD

[1]Psychology and Communication Technology Lab, Department of Psychology, Northumbria University, Newcastle upon Tyne, United Kingdom
[2]Dawes Centre for Future Crime, UCL Jill Dando Institute of Security and Crime Science, University College London, London, United Kingdom
[3]Department of Psychology, Northumbria University, Newcastle upon Tyne, United Kingdom

Corresponding Author:
Elizabeth Sillence, PhD
Psychology and Communication Technology Lab
Department of Psychology
Northumbria University
Northumberland Building
Newcastle upon Tyne
United Kingdom
Phone: 44 01912437246
Email: elizabeth.sillence@northumbria.ac.uk

Abstract

Background: The internet continues to offer new forms of support for health decision making. Government, charity, and commercial websites increasingly offer a platform for shared personal health experiences, and these are just some of the opportunities that have arisen in a largely unregulated arena. Understanding how people trust and act on this information has always been an important issue and remains so, particularly as the design practices of health websites continue to evolve and raise further concerns regarding their trustworthiness.

Objective: The aim of this study was to identify the key factors influencing US and UK citizens' trust and intention to act on advice found on health websites and to understand the role of patient experiences.

Methods: A total of 1123 users took part in an online survey (625 from the United States and 498 from the United Kingdom). They were asked to recall their previous visit to a health website. The online survey consisted of an updated general Web trust questionnaire to account for personal experiences plus questions assessing key factors associated with trust in health websites (information corroboration and coping perception) and intention to act. We performed principal component analysis (PCA), then explored the relationship between the factor structure and outcomes by testing the fit to the sampled data using structural equation modeling (SEM). We also explored the model fit across US and UK populations.

Results: PCA of the general Web trust questionnaire revealed 4 trust factors: (1) personal experiences, (2) credibility and impartiality, (3) privacy, and (4) familiarity. In the final SEM model, trust was found to have a significant direct effect on intention to act (beta=.59; P<.001), and of the trust factors, only credibility and impartiality had a significant direct effect on trust (beta=.79; P<.001). The impact of personal experiences on trust was mediated through information corroboration (beta=.06; P=.04). Variables specific to electronic health (eHealth; information corroboration and coping) were found to substantially improve the model fit, and differences in information corroboration were found between US and UK samples. The final model accounting for all factors achieved a good fit (goodness-of-fit index [0.95], adjusted goodness-of-fit index [0.93], root mean square error of approximation [0.50], and comparative fit index [0.98]) and explained 65% of the variance in trust and 41% of the variance in intention to act.

Conclusions: Credibility and impartiality continue to be key predictors of trust in eHealth websites. Websites with patient experiences can positively influence trust but only if users first corroborate the information through other sources. The need for corroboration was weaker in the United Kingdom, where website familiarity reduced the need to check information elsewhere. These findings are discussed in relation to existing trust models, patient experiences, and health literacy.

(J Med Internet Res 2019;21(11):e11125) doi:10.2196/11125

KEYWORDS

trust; eHealth; patient experiences

Introduction

Background

The number of people using the internet for health information and advice continues to grow with people affected by long-term or chronic conditions making particular use of online resources [1]. Over 80% of teens have sought health information online at some point about a range of health and lifestyle issues [2], and there has been a rise in surrogate seekers, those seeking information online for someone else [3]. Understanding how people come to trust the information and advice they find online has been an important issue since the widespread adoption of the internet [4] and continues to be so (see, eg, recent work by Marcu et al [5] and Lu et al [6]). The explosion in new providers, new formats, and platforms continues to generate concerns regarding the quality and variability of the health information available to the average citizen. Despite the introduction of codes and standards, for example, *Health on the Internet* code, early concerns over information quality, accuracy, and credibility [7] are still being echoed by researchers examining the provision of electronic health (eHealth) material across a range of conditions [8] including diabetes, osteoarthritis, and orthognathic surgery [9-11]. Today, such concerns sit within a wider debate about the veracity of information available to citizens through a variety of online sources. We know that people will often make snap judgments about the quality of information available online [4,12], relying upon simple heuristics to inform their decision making. We also know that people seldom make these judgments in isolation but are likely to show social influences in their information searches [12]. In particular, we can see that citizens exhibit homophily when going online for information—choosing to be guided by others they perceive as similar to themselves [13] and selecting information that is consistent with their own prior beliefs [14].

These social effects are particularly strong when people share their own health experiences online. Shared personal experiences are important to health consumers [15,16], and these are disseminated in online support communities, which can offer long-term supportive relationships, providing empathy, and reducing patients' sense of isolation [17,18]. As online social networks have grown, the range and availability of personal experiences have grown enormously. Peer-to-peer resources in the form of support forums, blogs, written or video testimonials, as well as curated experiences have become a common feature of online health resources. They are found in eHealth sites provided not just by concerned individuals but by charities, governmental organizations, and commercial websites alike. They once again put the concept of a trusting relationship center stage as the mediating technology, the host platform, and the contributors themselves can all be considered as objects of trust [19]. Put simply, a health consumer must typically make a number of layered trust decisions before engaging with peer-led material on a site [13], although a credible *host* site may be a prerequisite for trust in the more personal stories or blogs contained within [20].

The sheer number of eHealth resources available means that there are significant opportunities to check and verify any information and advice found online. Indeed, corroboration has been shown to have a key role in predicting trust and action around eHealth information [21]. However, research indicates that once again we see different factors influencing the layered trust decisions that are made [21]. At the peer-led level, when people are seeking to check information about personal experiences, these corroborating activities may become distorted by social networks, where the homophily effects of being able to tap into information bubbles of "people like me" may act to limit the effectiveness of cross-checking, particularly for groups with low socioeconomic status (eg, [14]). At the platform or website level, other factors come into play. Thus, for example, many websites require commercial funding, and this in turn can be signaled by the presence of online advertising, which in turn may act to undermine the perceived trustworthiness of the messages on the site. Impartiality is fundamental to trust in online resources [22], and advertising can lead a consumer to question the underlying motivations of an organization, sensing that they may not necessarily be acting primarily in the interests of a patient or carer [23]. Genuine peer contributors to a forum or site may wish to convey a credible, persuasive account of their experience with a particular product or service, but, if the narrative is framed in a commercial context, then the veracity of that experience may be called into question [24,25]. In general, it appears that personal experiences and commerce do not work well together. The blurring of the lines between testimonials and advertising serves to reduce the value of the personal accounts and the overall credibility of the website [23]. Furthermore, new trust concerns arise for both contributors and consumers of health content if people feel that the information they provide or access may be used to profile their own health status. This is a critical issue in the wake of new developments in the United States that give more freedom to internet service providers to sell on consumer information to advertisers [26], making health privacy a critical, but as yet under-researched aspect of trust decision making in eHealth [27].

Understanding the antecedents of trust in online health information has been a long-standing interest of the authors who, for the past 20 years, have developed and reported a number of large-scale eHealth surveys to gauge changes in the trust practices of people seeking health information online. Taken together, the studies have addressed the rise in patient-centered and patient-generated health information.

Since 2000, the range of patient-led resources and the nature and number of different eHealth providers have grown dramatically, and the most recent changes have seen a dramatic rise in patient narratives, often accompanied by new advertising funding models that may not always be viewed as appropriate in a health domain. The noticeable shift toward the inclusion of peer-led information creates interesting questions around what exactly it is that we are being asked to trust—the advice, the patient who provides a *story*, the organization behind the website, or other (sometimes unknown) funders. All of these can influence the decision to trust, and subsequently act upon, health advice. The extent to which *health privacy* affects trust in eHealth is also poorly understood. It is, therefore, timely to ask again about how people make their trust decisions.

Data we collected 10 years ago [21] resulted in a model that showed how trust in eHealth information and intention to act on the advice could be predicted on the basis of source credibility and impartiality. In that study, the predictive value of these 2 factors was enhanced when consumer responses to uncongenial health-risk information was taken into account. In particular, adding variables specific to health psychology (eg, measures addressing coping style), alongside measures designed to capture response to the online environment (eg, information corroboration), enhanced the model's predictive power.

Objectives

In this study, we aimed to update this model and provide a more timely understanding of the current antecedents of trust in online health information. We did this in 2 steps. First, we assessed the factorial structure of an updated general measure of trust in online health resources. We took the general measure of trust used in the study by Harris et al [21] and supplemented it with measures addressing inter alia personal experiences online, the presence of advertising, and health privacy concerns. Second, we sought to establish how well these subsequent factors improved the predictive power of the older model [21]. In addition, we purposely sampled from the United States and the United Kingdom to establish the robustness of the model across 2 widely different health care economies, one largely privatized (with funding via a complex health insurance network) and the other largely nationalized.

In summary, then, we sought to model the role of online personal experiences in health information and advice-seeking behavior using populations drawn from the United States and the United Kingdom.

Methods

Design

A cross-sectional survey was conducted in November 2015 and collected quantitative data from eHealth users regarding their use of health websites as part of a larger project measuring online trust in health websites every 5 years since 2000. We used a panel company to recruit a similar demographic to those that had participated in our previous studies to gain a sample representative enough to allow for meaningful comparisons.

Participants

A total of 8272 people clicked on the link from the recruitment advertisement on the panel company's internal Web page and were assessed on their eligibility to take part in the survey. Of this larger sample, 74.62% (6172/8272) indicated that they used the internet to look for health advice compared with 25.43% (2103/8272) who did not use the internet for advice. Following eligibility assessment (older than 18 years and UK- [40% quota] or US-based [60% quota]), a total of 1396 participants completed the questionnaire. A total of 96 were removed because of incomplete data resulting in 1123 participants that completed the full survey exploring online health seekers. Of the 1123 participants, 875 (77.92%) reported searching the internet for health advice for themselves, 145 (12.91%) for someone else, and the remaining 112 (9.97%) for both. Participants received £1.71 (or the US equivalent) for taking part in the study. Full details of participant demographics can be found in Table 1.

Table 1. Participant demographics (N=1123).

Participant characteristic	Online seekers frequency, n (%)
Location	
United States	625 (55.65)
United Kingdom	498 (44.34)
Gender	
Male	462 (41.14)
Female	661 (58.86)
Age (years)	
18-24	172 (15.32)
25-35	311 (27.69)
36-44	222 (19.77)
45-54	195 (17.36)
55-64	146 (13.00)
65+	63 (5.61)
Employment status	
Full time	545 (48.53)
Part time	171 (15.23)
Retired	137 (12.20)
Unemployed	208 (18.52)
Student	62 (5.52)
Marital status	
Single	354 (31.52)
Married	531 (47.28)
Cohabiting	106 (9.44)
Civil partnership	29 (2.58)
Divorced	84 (7.48)
Widowed	19 (1.69)
Ethnicity	
White	912 (81.21)
Latino/Hispanic	40 (3.56)
Middle Eastern	12 (1.07)
African	59 (5.25)
Caribbean	11 (0.98)
South Asian	23 (2.05)
East Asian	20 (1.78)
African American	11 (0.98)
Mixed	18 (1.60)
Prefer not to say	16 (1.42)
Highest level of education	
Less than high school/secondary school	18 (1.60)
Secondary school/high school/general educational development	294 (26.18)
Further education (college, A-levels or equivalent)	199 (17.72)
Bachelor's degree	490 (43.63)

Participant characteristic	Online seekers frequency, n (%)
Postgraduate degree (MSc, PhD)	122 (10.86)
Internet use (years)	
1-2	7 (0.62)
3-5	46 (4.10)
6-9	98 (8.73)
10-14	313 (27.87)
15-19	350 (31.17)
20+	309 (27.52)

Procedure

Before study commencement, the study received full ethical approval from the Department of Psychology at Northumbria University, and the online survey was piloted with 5 participants to assess comprehension and running of the survey. The survey was hosted on Qualtrics. The first page provided participants with information detailing the aim, length, data storage, contact details, and withdrawal process of study. They were then asked to provide informed consent. The study then commenced, and participants were asked whether they used the internet to look for health advice. Those answering "yes" then completed a series of questions relating to the last time they searched for health advice online. Specifically, they were asked to "think about any one site that you visited during that search" and to answer the remaining questions with respect to that site. They answered questions relating to the impact of the health advice on their coping perceptions and intention to act, the degree to which they trusted the information and website, and demographic information.

Measures

Unless stated otherwise, participants answered the following measures on a 5-point Likert scale (*1=strongly disagree* to *5=strongly agree*).

General Web Trust Questionnaire

The first measure contained the 24 items from the study by Harris et al [21], supplemented by 8 items assessing the presence of personal experiences [28] and 5 items to measure privacy concerns. In addition, *coping* was measured with 4 items such as "Looking at this site made me feel in control," in which participants' responses were rated on a 6-point scale with the following labels: *1=less, 2=slightly less, 3=no different, 4=slightly more, and 5=more* (Cronbach alpha=.83). *Information corroboration* with other sources of information was measured with the following 2 items: (1) "I checked other websites" and (2) "I checked other sources" (Cronbach alpha=.85). It is recognized that having just 2 items contributing to a measure can give challenge to the accuracy of Cronbach alpha, although in such cases alpha acts as a lower bound for the reliability, that is, it always underestimates the true reliability of the scale [29]. Note that these items were all taken from an earlier study [21].

Outcome Measures

Trust was measured with the following 2 items: (1) "I trusted the site" and (2) "I felt I could trust the information on the site" (Cronbach alpha=.78). *Intention to act* was an outcome measure, assessed with 1 item "I intended to act upon the advice."

Results

We first explored the updated general Web trust questionnaire by performing principal component analysis (PCA). We then explored the relationship between the factor structure and outcomes by testing its fit to the sampled data using structural equation modeling (SEM).

Properties of the General Web Trust Questionnaire

The 36 items of the scale were entered into the PCA, and varimax rotation with Kaiser normalization was used. Any items with factor loadings lower than 0.30 were suppressed (see Table 2).

The findings from the PCA revealed that 4 components (with eigenvalues above 1) could explain the data accounting for 66.057% of the variance. This complied with the minimum acceptable level of 60% variance and recommendations of eigenvalues above 1 for factors [30]. One item *The site was free from advertisements* did not load onto any component and was dropped from the analysis. In other words, this component was not, in isolation, a strong enough measure to be considered influential in the final model.

Overall, the analysis revealed that the 4 final components explained a large amount of the variance in the data, and the items had strong component loadings (well above the 0.30 criterion). It is recognized that the fourth component could be considered as weak as it only comprises 2 items. Advice is that there should be a minimum of 3 items per extracted component. However, it is reasonable that a component with 2 items is *identified*, provided that the other factors have more than 3 items and the 2-item factor has a nonzero covariance with at least one other factor in the population [31]. Such is the case here.

Table 2. Factor loadings for each item (factor loadings lower than 0.30 are suppressed).

Item	Rotation factor loadings			
	Factor 1: Personal experiences[a]	Factor 2: Credibility and impartiality[b]	Factor 3: Privacy[c]	Factor 4: Familiarity[d]
The language on the site made it easy to understand.	__[e]	0.783	—	—
The site helped me understand the issue better.	—	0.791	—	—
The site was easy to use.	—	0.780	—	—
The site told me most of what I needed to know.	—	0.692	—	—
The layout was consistent with other sites.	—	0.608	—	—
The advice appeared to be prepared by an expert.	—	0.664	—	—
The advice seemed to be offered in my best interests.	—	0.744	—	—
The advice came from a knowledgeable source.	—	0.714	—	—
The advice seemed credible.	—	0.747	—	—
The site was owned by a well-known organization.	—	—	—	0.769
The site featured familiar logos.	—	—	—	0.795
The site had a professional design.	—	0.679	—	—
The site had an attractive design.	—	0.605	—	—
The site provided reassurances about my privacy.	—	—	0.616	—
The site gave the option to post anonymously.	—	—	0.669	—
The site gave reassurances about how they used your information.	—	—	0.739	—
The site had a privacy policy.	—	—	0.717	—
The site explained their use of cookies.	—	—	0.637	—
The site contained accounts of other patient experiences.	0.815	—	—	—
There was a chance to share my experiences.	0.821	—	—	—
There were opportunities to interact with other people on the site.	0.829	—	—	—
On the site I saw a wide range of experiences rather different to mine.	0.791	—	—	—
The site offered powerful accounts of health experiences.	0.817	—	—	—
It felt like the advice was tailored to me personally.	0.559	—	—	—
On the site, I was offered the chance to see experiences from people just like me.	0.856	—	—	—
The site contained contributions from like-minded people.	0.863	—	—	—
I was able to contribute to content on the site.	0.817	—	—	—
The personal accounts on the site were written by people similar to me.	0.882	—	—	—
I found personal accounts that reflected my own experience.	0.875	—	—	—
I found personal accounts that were relevant to my condition.	0.876		—	—
There were opportunities to gather information from the personal accounts on the site.	0.870	—	—	—
The personal accounts contained advice for readers.	0.869	—	—	—
The personal accounts provided social or emotional support.	0.845	—	—	—
The advice appeared to be impartial and independent.	—	0.682	—	—
The advice seemed objective (ie, no hidden agenda).	—	0.695	—	—

Item	Rotation factor loadings			
	Factor 1: Personal experiences[a]	Factor 2: Credibility and impartiality[b]	Factor 3: Privacy[c]	Factor 4: Familiarity[d]
Removed item (the site was free from advertisements).	—	—	—	—

[a]Eigenvalue for factor 1 was 10.849, and the variance explained was 30.998%.

[b]Eigenvalue for factor 2 was 7.432, and the variance explained was 21.234%.

[c]Eigenvalue for factor 3 was 3.158, and the variance explained was 3.158%.

[d]Eigenvalue for factor 4 was 1.681, and the variance explained was 1.681%.

[e]Not applicable.

Structural Equation Modeling Analysis

The data were analyzed using SEM performed in IBM SPSS AMOS and based on the model structure from Harris et al [21], which represents the data collected 10 years ago. Maximum likelihood estimation methods were used to assess model fit, and the input for each analysis was the covariance matrix of the items. The goodness-of-fit for the models was evaluated with the following absolute goodness-of-fit indices (GFIs) [32]: (1) the Chi-square goodness-of-fit statistic; (2) the root mean square error of approximation (RMSEA); (3) GFI; (4) the adjusted goodness-of-fit (AGFI), and (5) comparative fit index (CFI). Nonsignificant Chi-square values indicate that the hypothesized model fits the data, and RMSEA values smaller than or equal to 0.08 are indicative of acceptable fit. However, values above 0.1 should lead to model rejection [33]. GFI values greater than 0.95 are indicative of good fit, and values greater than 0.90 are indicative of an acceptable fit [34]. AGFI values of 0.90 are indicative of a good fit, and values greater than 0.85 may be considered an acceptable fit [35]. The closer the CFI value is to 1 the better the fit [36].

The final model accounted for 65% of the variance in trust, 27% of the variance in coping, and 41% of the variance in intention to act. The model was a good fit for 4 of the indices. The fit indices for GFI and AGFI were 0.95 and 0.93, which are indicative of a good fit. RMSEA was 0.050, and CFI was 0.98. Path coefficients (beta) and R^2 values were also inspected in evaluating the predictive power of the models. Although the Chi-square indicated that the model was not a good fit to the data, $X^2_{168}=639.8$, $P<.001$, Chi-square has been criticized for being too sensitive to large sample sizes, especially for samples over 200 [37], as in this study.

Only credibility and impartiality was found to have a significant, direct relationship with trust (see Table 3). Familiarity and presence of personal experiences did not significantly relate to trust. The effects of familiarity, personal experiences, and privacy may be indirect and mediated through the other trust variables. In particular, personal experiences was found to have a significant direct effect on information corroboration, which in turn significantly predicted trust. Individuals who are presented with *personal experiences* may, therefore, corroborate this information with other sources and websites enhancing their trust in the personal experiences account. Trust in turn was found to significantly relate to coping perceptions and intention to act on the advice. This suggests that trustworthy websites heighten their coping perceptions, making them feel reassured, in control, and able to cope.

Table 3. The regression weights and critical ratio (ie, Z-score) values for the main effects of the hypothesized full model (combined UK and US participants).

Parameter	Unstandardized path coefficients	Critical ratio	*P* value
Credibility and impartiality → trust	0.944	17.110	<.001
Familiarity → trust	0.012	0.552	.58
PEX[a] → trust	0.021	0.960	.34
Information corroboration → trust	0.050	3.001	.003
Credibility and impartiality → information corroboration	0.520	7.566	<.001
Familiarity → information corroboration	−0.051	−1.289	.20
PEX → information corroboration	0.067	2.092	.04
Trust → coping	2.229	16.518	<.001
Trust → intention to act	0.794	16.197	<.001
Coping → intention to act	0.013	1.425	.15
Information corroboration → intention to act	0.063	2.751	.006

[a]PEX: personal experiences.

Comparison of Two Populations

A total of 2 further structural equation models were then assessed; one for each of the 2 populations that made up the full dataset, those from the United States and those from the United Kingdom. Although no previous literature exists to document consumer differences in terms of their trust in online health information, the countries differ widely in terms of state-run health provision, and it is known that health consumers differ in terms of their internet health behaviors [38] and that physicians in the United States and the United Kingdom differ widely in terms of their access to online information [39].

US Population

The model was a good fit for 4 of the indices. The GFI and AGFI were 0.93 and 0.91, respectively, and the RMSEA and CFI were 0.055 and 0.97, respectively, although the Chi-square indicated that the model was not a good fit to the data, X^2_{168}=481.3, *P*<.001 (see earlier above). Path coefficients (beta) and R^2 values were also inspected in evaluating the predictive power of the models. The final model accounted for 64% of the variance in trust, 27% of the variance in coping, and 44% of the variance in intention to act. Regression weights are presented in Table 4 below.

Table 4. The regression weights and critical ratio values for the main effects of the hypothesized model for US participants.

Parameter	Unstandardized path coefficients	Critical ratio	*P* value
Credibility and impartiality → trust	1.001	13.346	<.001
Familiarity → trust	–0.052	–1.515	.13
PEX[a] → trust	0.073	2.436	.02
Information corroboration → trust	0.068	3.023	.003
Credibility and impartiality → information corroboration	0.364	3.959	<.001
Familiarity → information corroboration	0.018	0.308	.76
PEX → information corroboration	0.060	1.408	.16
Trust → coping	2.224	12.696	<.001
Trust → intention to act	0.802	13.216	<.001
Coping → intention to act	0.008	0.651	.52
Information corroboration → intention to act	0.075	2.485	.01

[a]PEX: personal experiences.

There are 2 differences in the observed relationships when comparing the US model with the full model. First, the significant predictive path between personal experiences and information corroboration is lost. However, given that the regression weight is identical in both models, this is just a consequence of reduced power in the US analysis. More notable is the introduction of a significant path between personal experiences and trust that is not evident in the full model. All other paths are comparable between the 2 models.

UK Population

Although the Chi-square indicated that the model was not a good fit to the data, X^2_{168}=422.8, *P*<.001, the model was a good fit for the remaining 4 indices. The GFI and AGFI were 0.92 and 0.89, respectively. Finally, RMSEA was 0.055, and CFI was 0.97. Path coefficients (beta) and R^2 values were also inspected in evaluating the predictive power of the models. The final model accounted for 65% of the variance in trust, 27% of the variance in coping, and 38% of the variance in intention to act. Regression weights are presented in Table 5.

As with the US-based model, the significant predictive path between personal experiences and information corroboration is lost. Equally, however, the regression weight is identical in both models, and this is just a consequence of reduced power in the UK analysis. A total of 2 further paths also fail to reach significance in the UK model compared with the full model: information corroboration to trust and to intention to act. For these 2, there is a noticeable reduction in the regression coefficients for the UK model compared with both the US and full models, and as such the loss of significance is a consequence of a weaker relationship as well as a reduction in power. Moreover, the UK model also produces a significant path between familiarity and information corroboration that is not present in either the full or the US model.

In summary, although the US- and UK-based analyses share—as might be expected—many of the significant relationships identified in the full model, 2 distinct dissociations are also identified: The significant path between personal experiences and trust that only emerges in the US model and the significant (and negative) path between familiarity and information corroboration that only emerges in the UK model, such that UK citizens are less likely to corroborate information if their primary source is familiar.

Table 5. The regression weights and critical ratio values for the main effects of the hypothesized model for UK participants.

Parameter	Unstandardized path coefficients	Critical ratio	*P* value
Credibility and impartiality → trust	0.912	10.982	<.001
Familiarity → trust	0.034	1.135	.26
PEX[a] → trust	−0.031	−0.985	.33
Information corroboration → trust	0.029	1.141	.25
Credibility and impartiality → information corroboration	0.740	7.094	<.001
Familiarity → information corroboration	–0.139	–2.586	.01
PEX information corroboration	0.065	1.337	.18
Trust → coping	2.213	10.716	<.001
Trust → intention to act	0.782	9.929	<.001
Coping → intention to act	0.019	1.259	.21
Information corroboration → intention to act	0.058	1.656	.02

[a]PEX: personal experiences.

Discussion

Principal Findings

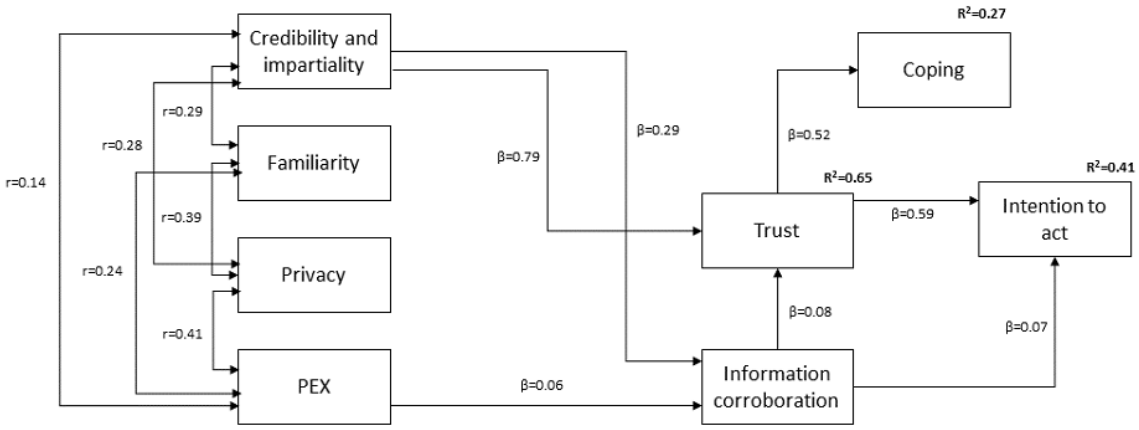
In terms of identifying the key predictors of trust and intention to act on health information, we found that trust significantly influenced self-reported intention to act on advice. Of the trust predictors, only credibility and impartiality was found to have a significant, direct relationship with trust. The effects of other variables (familiarity, personal experiences, and privacy) may be indirect and mediated through the other trust variables. For the role of personal experiences, it was found to have a significant direct effect on information corroboration, which in turn significantly predicted trust. Trust in turn was found to significantly relate to coping perceptions and intention to act on the advice. These results lead us to make the following observations.

The first point to note is that trust judgments significantly influence self-reported intention to act upon the health advice given online and furthermore, that these trust judgments reflect the extent to which people feel that the information sources are (1) credible, that is, contain good quality, relevant information, (2) well designed and presented, and (3) impartial, that is, contain information offered in the health consumer's best interest. These results resonate with recent findings in the existing literature (see Sbaffi and Rowley [40] for a systematic review). For example, trust is known to predict intention to act upon health advice (eg, [21,41]). The relevance, quality, usefulness, and accuracy of information are known determinants that the information content is trustworthy [42]. The presentation, ease of use, and clarity of information are linked to perceptions of *professionalism* that, again, underpin judgments of trust [4,43,44] and, finally, the beliefs about objectivity and impartiality of the source also ensure trust [21].

Looking in more detail at the model presented in Figure 1, we can see interesting similarities and differences between the current model and the model developed 10 years ago [21]. Specifically, Harris et al [21] showed that 2 website factors (information quality and impartiality) directly influenced trust. In the model we present here, these same 2 website factors, now combined into 1 construct named *credibility and impartiality* are the strongest predictors of trust. Harris et al [21] also showed that trust and its relationship to intention to act were moderated by 2 cognitive processes—involving threat appraisal and information corroboration. They also, along with personal experiences factor, significantly affect the processes of information corroboration, which in turn affect both trust and intention to act upon the health advice given, with a final *coping* factor also moderating the relationship between trust and intention to act. In short, credibility of information and impartiality occupy pivotal roles in our decision to trust the information we view online, as it did in the earlier Harris et al model [21], something entirely consistent with the ways in which patients come to develop trust relationships with their physician where there is a strong belief that doctors act in the patients' best interest (eg, [45]). Credibility and impartiality are key to trust in eHealth in 2019 as they were 10 years ago.

Online health information is also important in helping people to cope with health issues. When people trust sites that provide positive information about controlling symptoms or disease, it appears to help boost their overall sense of coping and efficacy. Although the model developed by Harris et al [21] did not find a significant relationship between trust and coping, we found that trust could account for 27% of the variance in coping. Our model confirms that placing trust in health websites is important in helping users to cope with their health issues, and this is in line with previous research indicating that seeking health information is in itself an important coping mechanism in enhancing adjustment to illness and in the promotion of health-related activities [46,47]. These findings could reflect a general improvement in the ways in which *trusted* sites offer information and advice and is possibly related to the rise of health websites offering patient experiences. We know from these data that patient experiences can influence trust (see below), and from the published literature, we know that personal experiences can also help people to feel supported in their health issues [18], but the ways that such experiences might directly affect coping would require further investigation.

Figure 1. The trust model with significant standardized path coefficients. PEX: personal experiences.



The role of personal experiences in relation to trust is an interesting finding, and our contribution here is novel. As we noted earlier, one of the biggest changes to the internet is the sharing of patient stories and experiences. We note in our model that personal experiences can influence trust but only indirectly through first, influencing those judgments of credibility and impartiality that are so important in predicting trust and second, influencing the ways that people choose to corroborate the information they view online. This finding resonates with the idea that although personal experiences are often liked, they are not necessarily trusted automatically [23]. The literature concerning trust in ecommerce and, in particular, social commerce provides a useful reference point for considering the relationship between trust in personal experiences and trust in the health website overall [48,49]. The trustworthiness of other customers on a website can be transferred to the community and thus help build stronger confidence or trust in the website as a whole [50]. Similarly, on social media sites, high levels of trust in other site members lead to higher levels of trust in and use of the site as a whole [51].

The corroboration point is interesting as in the combined UK and US data we found that low information credibility and impartiality, as well as the presence of personal experiences, led to higher levels of corroboration but that the need for corroboration, sometimes referred to as *triangulation*, differed between the US and UK samples. Specifically, in the United Kingdom, if the primary source of information was familiar, then patients expressed less need to corroborate that information. This could well be a function of the dominance of the National Health Service as a single trusted health care provider in the United Kingdom (indeed, most UK respondents cited the National Health Service website as their source of health information), as opposed to the more complex marriage of public and private insurance–based systems operating in the United States, where WebMD was the most popular online choice. It may also be possible that the difference could lie in the extent to which advertising was present in the most popular websites, but our single item on advertising was insufficient to provide good data here. These results do resonate with the data provided by Schneider et al [38] who compared eHealth search patterns in a private (United States) and public (United Kingdom) health care market and concluded that the US system incentivizes personal search into eHealth and that free access to health care professionals in the United Kingdom (including telephone support) reduces the incentive to search widely for health information online.

The health corroboration process relies upon people being able to make an appropriate distinction between the more or less reliable sources of information they find online, and of course people may differ in their ability to make this distinction and to retain it when trying to recall information at a later date. The extent to which individuals engage in information corroboration is likely to reflect eHealth literacy and suggest that we may need to think more carefully about how to support different individuals when making trust judgments about online health information [14,52]. This may be particularly important when personal experiences are present as we know that personal experiences can help trigger a homophily “patients like me” response that may mean individuals are yet more vulnerable to targeted messages [13].

Health privacy was introduced as a factor in this study. It did not impact directly on trust in eHealth information, but the effect of privacy may be indirect and mediated through other trust variables. The data for this study were collected before the introduction of new privacy and data protection legislation that regulates the storage of personal data. The General Data Protection Regulation in Europe, which came into force in May 2018, is designed to harmonize data protection law across Europe and to bring the law up to date with technological advancements, specifically the increasing use of digital data. It would be interesting to see how a more transparent and direct message about data processing may impact on people’s perceptions of data privacy with regard to health websites going forward, and in the wake of increasing public concerns about the privacy of their health data, it is interesting to note new models that speculate on the role of health privacy in eHealth [53].

Limitations and Future Work

Here we focus on a sample of the US and UK population, which limits how representative the findings are to other countries and cultures. Nevertheless, the model demonstrates the impact of trust in eHealth on health decision making for 2 different westernized countries (with different national health practices)

and where use of the internet and technology is widespread. However, further work is required to explore country and demographic differences such as the growing role of information credibility skills in navigating online information [54]. Second, future work would benefit from assessing the impact of advertising using a more comprehensive range of items.

We speculated that corroboration across online sources may be linked to advertising, but the single *advertising* item within the general trust questionnaire did not load onto any of the factors in the model. It may still be worth exploring this relationship in future work as it may point to a changing and increasingly complex situation concerning the form and presence of advertising on health websites. Advertising comes in many forms, from banners to embedded endorsements. Pharmaceutical sites offer a holistic form of advertising, although some sites may choose to advertise through the use of crafted personal testimonials. The single item assessing advertising may be too blunt an instrument to detect attitudes toward these different commercial approaches and limits what we can deduce about the effect of advertising on trust in eHealth. The work is underway to assess people's understanding of *advertising* more broadly in this context, especially given the blending of information sources in an online health care context [55].

Conclusions

In conclusion, despite the large increase in new providers, new formats, and platforms, impartiality continues to remain a key predictor of trust in health websites as well as the extent to which users consider information sources to be credible. The presence of personal experiences information can have a positive influence on trust provided that users corroborate the information through additional sources.

Conflicts of Interest

None declared.

References

1. Fox S, Duggan M, Purcell K. Pew Research Center. 2013 Jun 20. Family Caregivers are Wired for Health URL: https://www.pewinternet.org/2013/06/20/family-caregivers-are-wired-for-health/ [accessed 2019-10-24]
2. Wartella E, Rideout V, Montague H, Beaudoin-Ryan L, Lauricella A. Teens, health and technology: a national survey. Media Commun 2016;4(3):13. [doi: 10.17645/mac.v4i3.515]
3. Cutrona SL, Mazor KM, Vieux SN, Luger TM, Volkman JE, Finney Rutten LJ. Health information-seeking on behalf of others: characteristics of 'surrogate seekers'. J Cancer Educ 2015 Mar;30(1):12-19 [FREE Full text] [doi: 10.1007/s13187-014-0701-3] [Medline: 24989816]
4. Sillence E, Briggs P, Harris PR, Fishwick L. How do patients evaluate and make use of online health information? Soc Sci Med 2007 May;64(9):1853-1862. [doi: 10.1016/j.socscimed.2007.01.012] [Medline: 17328998]
5. Marcu A, Black G, Whitaker KL. Variations in trust in Dr Google when experiencing potential breast cancer symptoms: exploring motivations to seek health information online. Health Risk Soc 2018;20(7-8):325-341. [doi: 10.1080/13698575.2018.1550742]
6. Lu X, Zhang R, Wu W, Shang X, Liu M. Relationship between internet health information and patient compliance based on trust: empirical study. J Med Internet Res 2018 Aug 17;20(8):e253 [FREE Full text] [doi: 10.2196/jmir.9364] [Medline: 30120087]
7. Eysenbach G, Powell J, Kuss O, Sa E. Empirical studies assessing the quality of health information for consumers on the world wide web: a systematic review. J Am Med Assoc 2002;287(20):2691-2700. [doi: 10.1001/jama.287.20.2691] [Medline: 12020305]
8. Daraz L, Morrow AS, Ponce OJ, Farah W, Katabi A, Majzoub A, et al. Readability of online health information: a meta-narrative systematic review. Am J Med Qual 2018;33(5):487-492. [doi: 10.1177/1062860617751639] [Medline: 29345143]
9. Guan Y, Maloney KA, Roter DL, Pollin TI. Evaluation of the informational content, readability and comprehensibility of online health information on monogenic diabetes. J Genet Couns 2018 Jun;27(3):608-615 [FREE Full text] [doi: 10.1007/s10897-017-0155-y] [Medline: 28951986]
10. Barrow A, Palmer S, Thomas S, Guy S, Brotherton J, Dear L, et al. Quality of web-based information for osteoarthritis: a cross-sectional study. Physiotherapy 2018 Sep;104(3):318-326. [doi: 10.1016/j.physio.2018.02.003] [Medline: 30030036]
11. Lee KC, Berg ET, Jazayeri HE, Chuang S, Eisig SB. Online patient education materials for orthognathic surgery fail to meet readability and quality standards. J Oral Maxillofac Surg 2019 Jan;77(1):180.e1-180.e8. [doi: 10.1016/j.joms.2018.08.033] [Medline: 30296407]
12. Metzger MJ, Flanagan AJ, Medders RB. Social and heuristic approaches to credibility evaluation online. J Commun 2010;60(3):413-439. [doi: 10.1111/j.1460-2466.2010.01488.x]
13. Sillence E, Hardy C, Harris PR, Briggs P. Modeling Patient Engagement in Peer-to-Peer Healthcare. In: Proceedings of the 23rd International Conference on World Wide Web. 2014 Presented at: WWW'14 Companion; April 7-11, 2014; Seoul, Korea p. 481-486. [doi: 10.1145/2567948.2576937]

14. Perez SL, Kravitz RL, Bell RA, Chan MS, Paterniti DA. Characterizing internet health information seeking strategies by socioeconomic status: a mixed methods approach. BMC Med Inform Decis Mak 2016 Aug 9;16:107 [FREE Full text] [doi: 10.1186/s12911-016-0344-x] [Medline: 27506607]
15. Lovatt M, Bath PA, Ellis J. Development of trust in an online breast cancer forum: a qualitative study. J Med Internet Res 2017 May 23;19(5):e175 [FREE Full text] [doi: 10.2196/jmir.7471] [Medline: 28536093]
16. Brady E, Segar J, Sanders C. 'You get to know the people and whether they're talking sense or not': negotiating trust on health-related forums. Soc Sci Med 2016 Aug;162:151-157 [FREE Full text] [doi: 10.1016/j.socscimed.2016.06.029] [Medline: 27359321]
17. Rozmovits L, Ziebland S. What do patients with prostate or breast cancer want from an Internet site? A qualitative study of information needs. Patient Educ Couns 2004 May;53(1):57-64. [doi: 10.1016/S0738-3991(03)00116-2] [Medline: 15062905]
18. Ziebland S, Wyke S. Health and illness in a connected world: how might sharing experiences on the internet affect people's health? Milbank Q 2012 Jul;90(2):219-249 [FREE Full text] [doi: 10.1111/j.1468-0009.2012.00662.x] [Medline: 22709387]
19. Grabner-Kräuter S. Web 2.0 social networks: the role of trust. J Bus Ethics 2010 Aug;90(S4):505-522. [doi: 10.1007/s10551-010-0603-1]
20. Hu Y, Shyam Sundar S. Effects of online health sources on credibility and behavioral intentions. Commun Res 2010;37(1):105-132. [doi: 10.1177/0093650209351512]
21. Harris PR, Sillence E, Briggs P. Perceived threat and corroboration: key factors that improve a predictive model of trust in internet-based health information and advice. J Med Internet Res 2011 Jul 27;13(3):e51 [FREE Full text] [doi: 10.2196/jmir.1821] [Medline: 21795237]
22. Powell J, Inglis N, Ronnie J, Large S. The characteristics and motivations of online health information seekers: cross-sectional survey and qualitative interview study. J Med Internet Res 2011 Mar 23;13(1):e20 [FREE Full text] [doi: 10.2196/jmir.1600] [Medline: 21345783]
23. Sillence E, Hardy C, Briggs P. Why Don't We Trust Health Websites That Help Us Help Each Other?: An Analysis of Online Peer-to-Peer Healthcare. In: Proceedings of the 5th Annual ACM Web Science Conference. 2013 Presented at: WebSci'13; May 2-4, 2013; Paris, France p. 396-404. [doi: 10.1145/2464464.2464488]
24. Vance K, Howe W, Dellavalle RP. Social internet sites as a source of public health information. Dermatol Clin 2009 May;27(2):133-6, vi. [doi: 10.1016/j.det.2008.11.010] [Medline: 19254656]
25. Hohm C, Snyder J. 'It Was the Best Decision of My Life': a thematic content analysis of former medical tourists' patient testimonials. BMC Med Ethics 2015 Jan 22;16:8 [FREE Full text] [doi: 10.1186/1472-6939-16-8] [Medline: 25614083]
26. Brewster T. Forbes. 2017 Mar 30. Now Those Privacy Rules Are Gone, This Is How ISPs Will Actually Sell Your Personal Data URL: https://www.forbes.com/sites/thomasbrewster/2017/03/30/fcc-privacy-rules-how-isps-will-actually-sell-your-data/#741fe40e21d1 [accessed 2019-10-24]
27. Kim Y. Trust in health information websites: a systematic literature review on the antecedents of trust. Health Informatics J 2016 Jun;22(2):355-369. [doi: 10.1177/1460458214559432] [Medline: 25518944]
28. Sillence L, Mo P, Briggs P, Harris PR. The Changing Face of Trust in Health Websites. SSRN Journal 2011:-. [doi: 10.2139/ssrn.1920317]
29. Eisinga R, Grotenhuis MT, Pelzer B. The reliability of a two-item scale: Pearson, Cronbach, or Spearman-Brown? Int J Public Health 2013 Aug;58(4):637-642. [doi: 10.1007/s00038-012-0416-3] [Medline: 23089674]
30. Hinkin TR. A brief tutorial on the development of measures for use in survey questionnaires. Organ Res Methods 1998;1(1):104-121. [doi: 10.1177/109442819800100106]
31. Yong AG, Pearce S. A beginner’s guide to factor analysis: focusing on exploratory factor analysis. Tutor Quant Methods Psychol 2013;9(2):79-94. [doi: 10.20982/tqmp.09.2.p079]
32. Jöreskog KG, Sörbom D. LISREL VI: Analysis of Linear Structural Relationships by Maximum Likelihood, Instrumental Variables, and Least Squares Methods, Users Guide. Mooresville, Indiana: Scientific Software; 1986.
33. Browne MW, Cudeck R. Alternative ways of assessing model fit. Soc Methods Res 1992;21(2):230-258. [doi: 10.1177/0049124192021002005]
34. Marsh H, Grayson D. Latent variable models of multitrait-multimethod data. In: Hoyle R, editor. Structural Equation Modeling: Concepts, Issues, and Applications. Thousand Oaks, CA: Sage; 1995:177-198.
35. Hu L, Bentler P. Evaluating model fit. In: Hoyle R, editor. Structural Equation Modeling: Concepts, Issues, and Applications. Thousand Oaks, CA: Sage; 1995:76-99.
36. Garson GD. Faculty and Staff - NC State. 2009. Statnotes: Topics in Multivariate Analysis URL: https://faculty.chass.ncsu.edu/garson/PA765/statnote.htm [accessed 2019-10-31]
37. Hoe S. Issues and procedures in adopting structural equation modeling technique. J Appl Quant Methods 2008;3(1):76-83 [FREE Full text]
38. Schneider JA, Holland CP. eHealth search patterns: a comparison of private and public health care markets using online panel data. J Med Internet Res 2017 Apr 13;19(4):e117 [FREE Full text] [doi: 10.2196/jmir.6739] [Medline: 28408362]
39. Davies KS. Physicians and their use of information: a survey comparison between the United States, Canada, and the United Kingdom. J Med Libr Assoc 2011 Jan;99(1):88-91 [FREE Full text] [doi: 10.3163/1536-5050.99.1.015] [Medline: 21243061]

40. Sbaffi L, Rowley J. Trust and credibility in web-based health information: a review and agenda for future research. J Med Internet Res 2017 Jun 19;19(6):e218 [FREE Full text] [doi: 10.2196/jmir.7579] [Medline: 28630033]
41. Sheng X, Simpson PM. Health care information seeking and seniors: determinants of Internet use. Health Mark Q 2015;32(1):96-112. [doi: 10.1080/07359683.2015.1000758] [Medline: 25751321]
42. Corritore CL, Wiedenbeck S, Kracher B, Marble RP. Online trust health information websites. Int J Technol Hum Interact 2012;8(4):92-115. [doi: 10.4018/jthi.2012100106]
43. Johnson F, Rowley J, Sbaffi L. Modelling trust formation in health information contexts. J Inf Sci 2015;41(4):415-429. [doi: 10.1177/0165551515577914]
44. Fisher J, Burstein F, Lynch K, Lazarenko K. 'Usability + usefulness = trust': an exploratory study of Australian health web sites. Internet Res 2008 Oct;18(5):477-498. [doi: 10.1108/10662240810912747]
45. Hall MA, Zheng B, Dugan E, Camacho F, Kidd KE, Mishra A, et al. Measuring patients' trust in their primary care providers. Med Care Res Rev 2002 Oct;59(3):293-318. [doi: 10.1177/1077558702059003004] [Medline: 12205830]
46. Medlock S, Eslami S, Askari M, Arts DL, Sent D, de Rooij SE, et al. Health information-seeking behavior of seniors who use the internet: a survey. J Med Internet Res 2015 Jan 8;17(1):e10 [FREE Full text] [doi: 10.2196/jmir.3749] [Medline: 25574815]
47. Lambert SD, Loiselle CG. Health information seeking behavior. Qual Health Res 2007 Oct;17(8):1006-1019. [doi: 10.1177/1049732307305199] [Medline: 17928475]
48. Kim DJ. Self-perception-based versus transference-based trust determinants in computer-mediated transactions: a cross-cultural comparison study. J Manag Inf Syst 2008;24(4):13-45. [doi: 10.2753/mis0742-1222240401]
49. Chow WS, Shi S. Understanding Consumer Trust in Social Commerce Websites. In: Proceedings of 18th Pacific Asia Conference on Information Systems. 2014 Presented at: PACIS'14; June 24-28, 2014; Chengdu, China.
50. Ng CS. Intention to purchase on social commerce websites across cultures: a cross-regional study. Inf Manag 2013;50(8):609-620. [doi: 10.1016/j.im.2013.08.002]
51. Valenzuela S, Park N, Kee KF. Is there social capital in a social network site?: Facebook use and college students' life satisfaction, trust, and participation. J Comput Mediat Commun 2009;14(4):875-901. [doi: 10.1111/j.1083-6101.2009.01474.x]
52. Paige SR, Krieger JL, Stellefson ML. The influence of eHealth literacy on perceived trust in online health communication channels and sources. J Health Commun 2017 Jan;22(1):53-65 [FREE Full text] [doi: 10.1080/10810730.2016.1250846] [Medline: 28001489]
53. Shen N, Strauss J, Silver M, Carter-Langford A, Wiljer D. The eHealth trust model: a patient privacy research framework. In: Lau F, Bartle-Clar JA, Bliss J, Borycki EM, Courtney KL, editors. Improving Usability, Safety and Patient Outcomes with Health Information Technology. Amsterdam: IOS Press; 2019:382-387.
54. Armstrong-Heimsoth A, Johnson ML, McCulley A, Basinger M, Maki K, Davison D. Good Googling: a consumer health literacy program empowering parents to find quality health information online. J Consum Health Internet 2017;21(2):111-124. [doi: 10.1080/15398285.2017.1308191]
55. Sullivan HW, O'Donoghue AC, Rupert DJ, Willoughby JF, Amoozegar JB, Aikin KJ. Are disease awareness links on prescription drug websites misleading? A randomized study. J Health Commun 2016 Nov;21(11):1198-1207. [doi: 10.1080/10810730.2016.1237594] [Medline: 27805473]

Abbreviations

AGFI: adjusted goodness-of-fit
CFI: comparative fit index
eHealth: electronic health
GFI: goodness-of-fit index
PCA: principal component analysis
RMSEA: root mean square error of approximation
SEM: structural equation modeling

Edited by G Eysenbach; submitted 24.05.18; peer-reviewed by L Sbaffi, K Aikin, J Villegas; comments to author 05.10.18; revised version received 28.02.19; accepted 02.09.19; published 11.11.19.

Please cite as:
Sillence E, Blythe JM, Briggs P, Moss M
A Revised Model of Trust in Internet-Based Health Information and Advice: Cross-Sectional Questionnaire Study
J Med Internet Res 2019;21(11):e11125
URL: http://www.jmir.org/2019/11/e11125/
doi:10.2196/11125
PMID:31710297

Original Paper

Barriers and Facilitators to the Implementation of eHealth Services: Systematic Literature Analysis

Björn Schreiweis[1,2], Dr sc hum; Monika Pobiruchin[2,3], Dr sc hum; Veronika Strotbaum[2,4], MA; Julian Suleder[2,5], MSc; Martin Wiesner[2,6], Dipl-Inform Med; Björn Bergh[1], Dr med

[1]Institute for Medical Informatics and Statistics, University Hospital Schleswig-Holstein and Kiel University, Kiel, Germany

[2]Consumer Health Informatics Special Interest Group, German Association for Medical Informatics, Biometry and Epidemiology eV, Cologne, Germany

[3]GECKO Institute for Medicine, Informatics and Economics, Heilbronn University, Heilbronn, Germany

[4]Zentrum für Telematik und Telemedizin GmbH, Bochum, Germany

[5]ERNW Research GmbH, Heidelberg, Germany

[6]Department of Medical Informatics, Heilbronn University, Heilbronn, Germany

Corresponding Author:
Björn Schreiweis, Dr sc hum
Institute for Medical Informatics and Statistics
University Hospital Schleswig-Holstein and Kiel University
Arnold-Heller-Straße 3
Kiel, 24105
Germany
Phone: 49 431500 ext 30701
Fax: 49 43150030704
Email: bjoern.schreiweis@uksh.de

Abstract

Background: The field of eHealth has a history of more than 20 years. During that time, many different eHealth services were developed. However, factors influencing the adoption of such services were seldom the main focus of analyses. For this reason, organizations adopting and implementing eHealth services seem not to be fully aware of the barriers and facilitators influencing the integration of eHealth services into routine care.

Objective: The objective of this work is to provide (1) a comprehensive list of relevant barriers to be considered and (2) a list of facilitators or success factors to help in planning and implementing successful eHealth services.

Methods: For this study, a twofold approach was applied. First, we gathered experts' current opinions on facilitators and barriers in implementing eHealth services via expert discussions at two health informatics conferences held in Europe. Second, we conducted a systematic literature analysis concerning the barriers and facilitators for the implementation of eHealth services. Finally, we merged the results of the expert discussions with those of the systematic literature analysis.

Results: Both expert discussions (23 and 10 experts, respectively) identified 15 barriers and 31 facilitators, whereas 76 barriers and 268 facilitators were found in 38 of the initial 56 articles published from 12 different countries. For the analyzed publications, the count of distinct barriers reported ranged from 0 to 40 (mean 10.24, SD 8.87, median 8). Likewise, between 0 and 48 facilitators were mentioned in the literature (mean 9.18, SD 9.33, median 6). The combination of both sources resulted in 77 barriers and 292 facilitators for the adoption and implementation of eHealth services.

Conclusions: This work contributes a comprehensive list of barriers and facilitators for the implementation and adoption of eHealth services. Addressing barriers early, and leveraging facilitators during the implementation, can help create eHealth services that better meet the needs of users and provide higher benefits for patients and caregivers.

(J Med Internet Res 2019;21(11):e14197) doi:10.2196/14197

KEYWORDS

eHealth; health information interoperability; policy; software design

Introduction

Background

In 1999, the term *eHealth* was coined. The first publications defined it as a "new term needed to describe the combined use of electronic communication and information technology in the health sector. The use in the health sector of digital data—transmitted, stored, and retrieved electronically—for clinical, educational, and administrative purposes, both at the local site and at a distance" [1].

With the evolution of an increasing number of e-services, health care is providing many different eHealth services. In general, eHealth is associated with a positive influence on health care outcomes [2]. Improved cost-effectiveness, more information on a patient's health status, and better communication between health care professionals are just some examples of the benefits of eHealth services [3,4]. However, there is no consistent picture of eHealth services' adoption and broad acceptance. Often, eHealth services are not adopted and lack acceptance by their users [5]. However, for several services, domains, or patient groups, the levels of acceptance and related adoption rates are reported to be higher [6].

For several years, health care institutions have evaluated and started to use eHealth services to support patient care. The evolution of mobile phones and the broad availability of apps for prevention, wellness, and fitness scenarios has resulted in an increased importance of eHealth for the health care industry because these new services help to better support care processes [7,8]. Both the primary and secondary health care markets are important when it comes to eHealth services. With eHealth being an important economic factor, member countries of the Organisation for Economic Co-operation and Development spend an average of 8.9% of their gross domestic product on health [9]. In addition, start-ups and "big players" (eg, Google, Apple, Facebook, Amazon, and Microsoft) [10] play an important role in the eHealth economy. These economic factors drive changes in eHealth legislation in national health care systems, such as the eHealth Act in Germany and the Electronic Patient Record Act in Switzerland [11].

Several models are available to evaluate the use of technology (eg, Technology Acceptance Model [12] and Unified Theory of Acceptance and Use of Technology [13]), which are often adapted for evaluation in the eHealth domain [14,15]. Such models provide criteria for the evaluation of technology acceptance. The use of eHealth in routine care can be explained to a certain extent by these models [14-16]. However, those models may benefit from several additions and modifications, especially in relation to implementation.

Several reviews and projects have identified barriers and facilitators for eHealth service adoption in certain environments and disease contexts, such as mental health [17-19], veterans health care [20-22], and hypertension [23-25]. However, to the best of our knowledge, no overview, meta-analysis, or comprehensive list of barriers and facilitators affecting the adoption of eHealth services has been conducted and published. For this reason, we organized two expert workshops and related discussion rounds to obtain an overview of the barriers and facilitators for the adoption of eHealth services. Both workshops were independent of a specific scenario and were accompanied by an exhaustive literature analysis.

Objective

The objective of this work is twofold: to provide (1) a comprehensive list of barriers to be considered, and (2) a list of facilitators or success factors to help in planning and implementing eHealth services. It is not within the scope of this paper to provide another model for evaluating eHealth or telemedicine services.

Methods

Overview

Two different approaches were combined in this study. First, we wanted to obtain international experts' current opinions on facilitators and barriers toward the implementation of eHealth services. This step helped to identify immediate experiences and knowledge bases especially from experts from countries with a higher level of digitization in health care. Second, we were interested in facilitators and barriers for implementation in completed projects and initiatives. Thus, two rounds of expert discussions at health informatics conferences in Europe were organized and held. Finally, a systematic literature analysis on barriers and facilitators for the implementation and adoption of eHealth applications was conducted.

Expert Discussions

Two expert discussions were organized at conferences in Europe: (1) Medical Informatics Europe (MIE) 2015 in Madrid, Spain, and (2) eHealth Innovation Days (eHID) 2017 in Flensburg, Germany.

The MIE 2015 expert discussion in Madrid included 23 international experts from the field of medical and health informatics (15 participants from Europe; 5 from the Middle East, Asia, and America; and 3 German organizers). The primary topic of the expert discussion was "Consumer Health Informatics: Barriers and Facilitators of eHealth Usage Among Consumers." The discussion included three short introduction talks, followed by three discussion groups on barriers for eHealth use among consumers [26]. However, due to time constraints resulting from the workshop format, the discussion mostly focused on barriers. Each group separately discussed the barriers to the use of eHealth applications and wrote them down on prompt cards. Once the discussions of five small groups were finished, each group presented their results briefly. The organizers of the expert discussion collected and aggregated the results in the format of a short workshop report [26].

The second expert discussion, on the topic of "Success Factors for Consumer-Centered eHealth Services," was held in Flensburg, Germany. Participants were experts in the fields of medical and health informatics located in the Baltic Sea region, especially Sweden, Finland, Estonia, and Germany. An introduction followed by two short keynote talks constituted a starting point for the experts changing their perspective to one

of five stakeholder groups. There were stakeholder groups for (1) citizens, patients, and family members (3 experts); (2) start-ups and application developers (4 experts); (3) researchers (3 experts); (4) policy makers and politicians (0 experts); and (5) data privacy officers and chief information officers (CIOs) (0 experts). The stakeholder groups for policy makers and politicians, as well as data privacy officers and CIOs were planned but were called off (0 participants). Each stakeholder group brainstormed on the success factors and facilitators for consumer-centric eHealth application use and/or its implementation. Next, each group presented briefly, and all groups discussed the results in a panel format. The results of each group were collected via flipcharts and consolidated by the authors in similar formats as the results of the first expert workshop held during MIE 2015.

Literature Analysis

To identify relevant articles in the field, a PubMed search was conducted on May 28, 2018, including the following query terms: ((“telemedicine”[MeSH Terms] OR “telemedicine”[All Fields] OR “ehealth”[All Fields]) AND (“adoption”[MeSH Terms] OR “adoption”[All Fields]) AND barriers[All Fields] AND facilitators[All Fields]) AND ((“patients”[MeSH Terms] OR “patients”[All Fields]) OR consumers[All Fields]).

The time frame for potentially relevant articles was only limited by the search date. All articles published before this retrieval date were considered relevant. The resulting literature was filtered by scanning for actual mentions of barriers or facilitators for the adoption or implementation of any kind of eHealth application (see Textbox 1). In this context, titles, abstracts, and full-text articles were read to determine whether the article met the aforementioned criteria. For all identified papers, barriers and facilitators were extracted manually by one of the authors. Barriers and facilitators were listed in an Excel spreadsheet (see Multimedia Appendix 1). Next, a categorization was applied creating a mind map for barriers and facilitators separately (see Multimedia Appendix 2). This categorization was based on the three main categories as identified by Griebel et al [26]: (1) individual, (2) environmental and organizational, and (3) technical.

Textbox 1. Criteria for the inclusion and exclusion criteria of literature in the analysis.

Inclusion criteria

- Published and listed on PubMed as of May 28, 2018
- Listing barriers for the implementation or adoption of eHealth services and/or listing success factors/facilitators for the implementation or adoption of eHealth services
- Articles in English and German

Exclusion criteria

- Article about research protocols of a planned study (ie, no results on barriers and/or facilitators)
- Abstract not available
- Article not about eHealth services
- Full text not accessible

Comparison of Expert Discussions and Literature Analysis

One expert in the field of medical informatics categorized the barriers identified in the literature according to the categories provided by the study of Griebel et al [26] and extended the original mind map with the results from the literature analysis conducted for this study. The success factors for eHealth service adoption identified in the literature were categorized using the main categories (individual, environmental and organizational, and technical) in accordance with the categorization of barriers. The subcategorization of the results of the expert discussions refining the three categories was done where applicable and subsequently reviewed by a coauthor. The mind map, originally generated with results from the expert discussion on success factors, was then augmented with items found in the literature. Finally, the aggregated results of the Griebel et al study [26] were extended with the results from the expert workshop on success factors, with facilitators found in the literature, and displayed in a hierarchical form (mind map).

Results

Results of the findings of the expert discussions are outlined first, followed by the results from the literature review. Both result sets are then compared for common and different attributes.

Expert Discussions

The expert discussion concerning barriers for eHealth services resulted in three categories of barriers: (1) individual, (2) environmental and organizational, and (3) technical barriers (see Figure 1). The category of individual barriers aggregated cognitive, motivational, accessibility, and trust-related barriers of individual consumers. Financial issues, political barriers, and organizational structures formed the category of environmental and organizational barriers. Unsuited services or design not fitting to the users’ needs were among the technical barriers. Security concerns were another barrier because often systems and network-enabled medical devices fail to provide an acceptable level of security. Additionally, system language, missing support (who to call for help?), missing standards (both for patient data and for data exchange), and missing system

feedback leading to unclear benefits were mentioned as barriers for eHealth services.

The expert discussion focusing on success factors and facilitators of consumer-centric eHealth services resulted in similar categories (individual success factors, environmental success factors, and technical success factors) (see Figure 2). We identified 31 success factors in the expert discussions. Subcategories of the individual success factors were a clear benefit of the service, trusting and controlling the service, the collaboration via the service, the service's user experience, and that the service facilitates research. Flexible funding, health outcomes, policies for using generated data for research, competition, and supporting laws and regulations were the subcategories of environmental success factors. Usability, standards, security, and reliability of the service were subcategories of technical success factors.

Figure 1. Barriers of eHealth usage among consumers identified in the first expert discussion at MIE 2015.

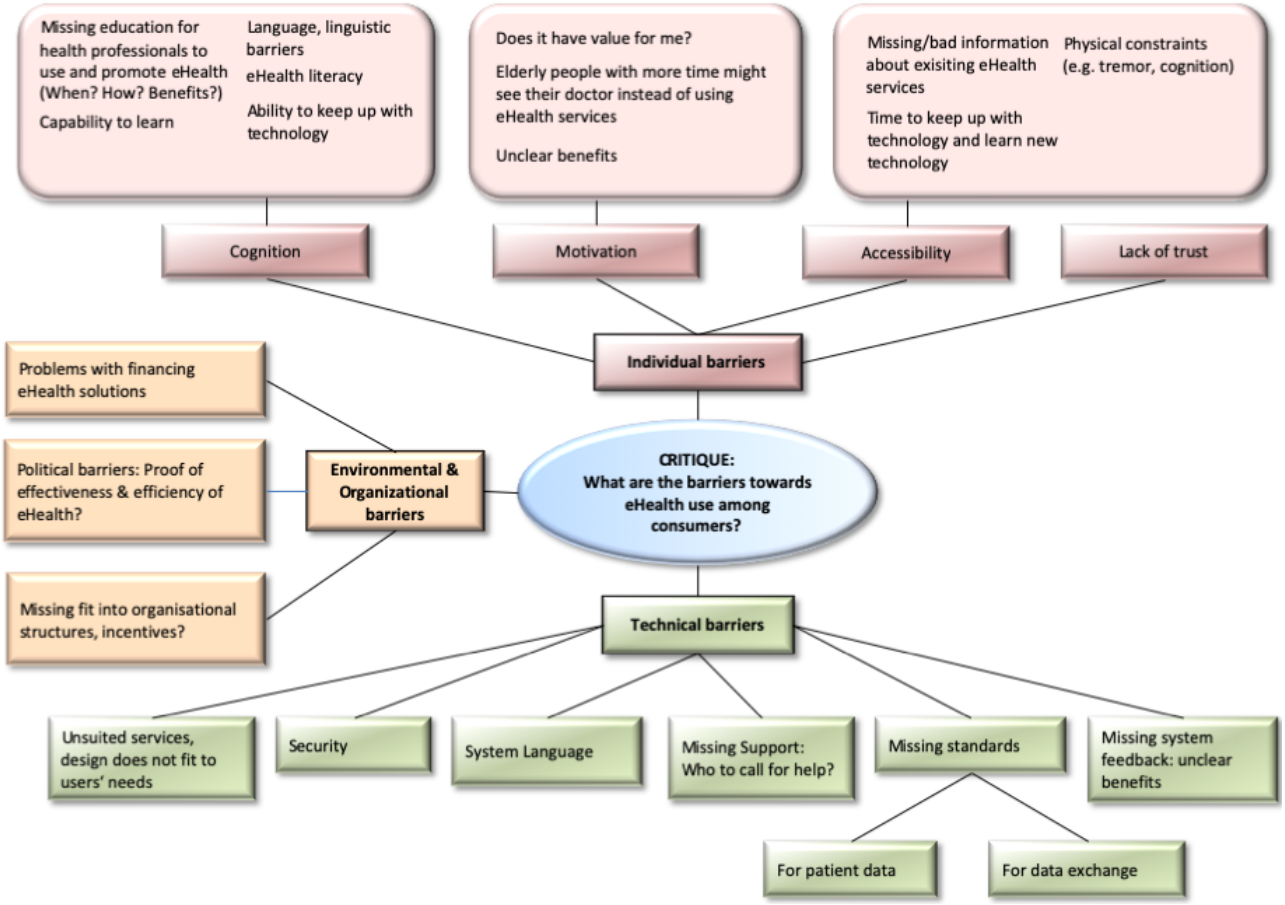
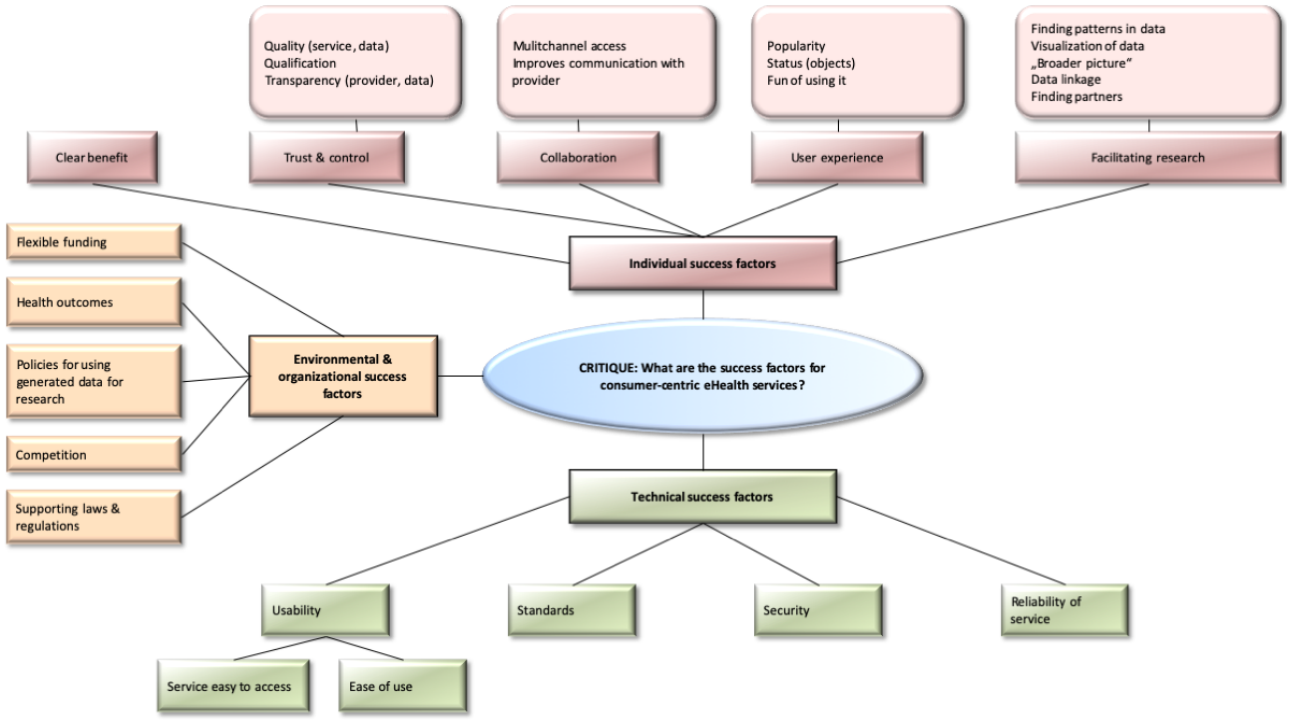


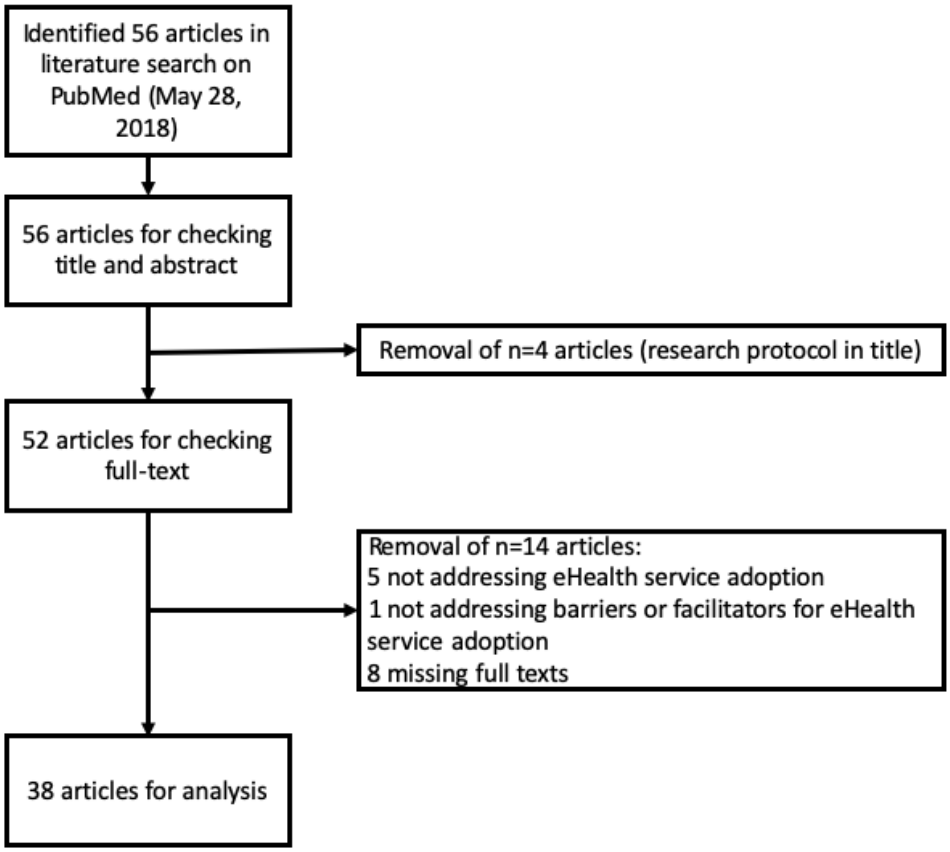
Figure 2. Success factors for consumer-centric eHealth services identified in the second expert discussion at eHID.



Literature Analysis

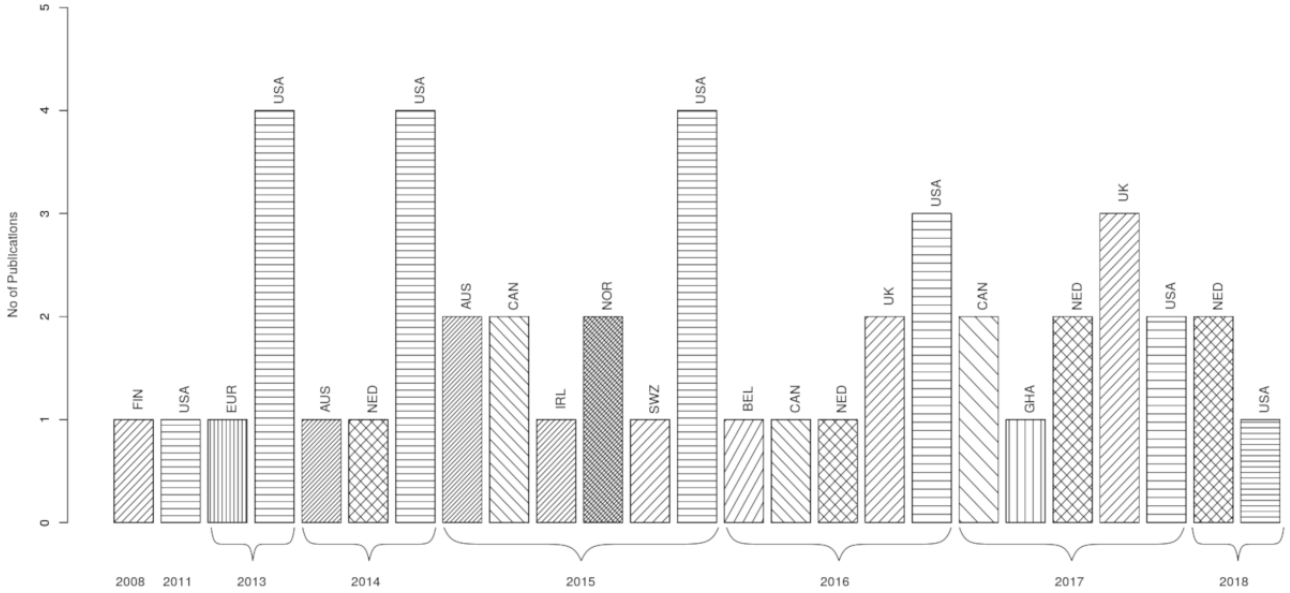
The literature analysis resulted in 56 publications published between December 27, 2007, and May 3, 2018. Of these publications, 38 were found to be relevant with full texts accessible to the authors for in-depth analyses [17-25,27-55] (see Figure 3). For the excluded 18 publications, either the full text was not accessible to the authors (n=8) or the articles did not describe, analyze, or present results about barriers or facilitators for the use of eHealth applications (n=10) (exclusion criteria see Textbox 1).

Figure 3. Flowchart for the identification of articles meeting the inclusion or exclusion criteria (see Textbox 1).



Publications, including the ones missing full text (n=8), originated mostly from the United States (19/46, 41%, published 2011-2018), followed by the Netherlands (6/46, 13%, published 2014-2018), Canada (5/46, 10%, published 2015-2017), the United Kingdom (5/46, 10%, published 2016-2017), Australia (3/46, 6%, published 2014-2015), and Norway (2/46, 4.35%, published 2015). Ghana (published 2017), Belgium (published 2016), Ireland (published 2015), Swaziland (published 2015), Europe (published 2013), and Finland (published 2008) each had one publication in this literature analysis (1/46, 2%; see Figure 4).

Figure 4. Included publications (before removal of missing full texts) by year and geographical location.



We identified 76 distinct barriers (33 individual, 25 environmental, and 18 technical) and a total of 268 facilitators (131 individual, 101 environmental, and 36 technical) in the literature (see Multimedia Appendix 1). The most frequent barrier in the literature was limited exposure/knowledge of eHealth (ie, poor digital health literacy) with 16 references [17,19-22,32,33,36,38,40,41,43,47,48,51,55], followed by 15 references of lack of necessary devices [19,20,22,24,32,37,38,40,41,43,44,48,51,53,55], and problems with financing eHealth solutions [17-19,21,24,27,28, 33,38,39,42,47]. For a complete list of the top 10 barriers, see Table 1.

Ease of use was the most stated success factor found in the literature with seven references [20,30,33,43,44,47,53], followed by improves communication [23,28,35,44,55]. For a list of the top six facilitators, see Table 2.

Table 1. List of top 10 barriers mentioned in the literature.

Position	Perceived barrier	Mentions, n	References
1	Limited exposure/knowledge of eHealth (eg, poor digital health literacy)	16	[17,19-22,32,33,36,38,40,41,43,47,48,51,55]
2	Lack of necessary devices	15	[19,20,22,24,32,37,38,40,41,43,44,48,51,53,55]
2	Problems with financing eHealth solutions	15	[17-19,21,24,27,28,33,38,39,42,47]
4	Cognition	13	[19,23,24,30,36-38,44,47,51-54]
4	Security	13	[17,24,28,32-35,38-40,43,51,55]
6	Motivation	12	[17,23,24,36,39-41,43-45,53,55]
7	Accessibility	10	[23,36,38,39,43,46,47,51,53,55]
8	Unsuited services, design does not fit users' needs	9	[24,33,35,38,43,46,47,52,53]
8	Confidentiality	9	[17,23,30-34,40,51]
10	Missing fit into organizational structures, incentives	8	[18,24,25,27,28,42,45,47]
10	Added workload	8	[23,24,28,29,38,39,44,54]

Table 2. List of the top six facilitators mentioned in literature[a].

Position	Perceived facilitator	Mentions, n	References
1	Ease of use	7	[20,30,33,43,44,47,53]
2	Improves communication	5	[23,28,35,44,55]
2	Motivation	5	[23,36,47,52,53]
2	Integrated into care	5	[24,25,28,33,35]
5	Involvement of all relevant stakeholders	4	[18,25,47,54]
6	Availability of resources	3	[19,25,54]
6	User-friendliness	3	[27,36,39]

[a] There were too many facilitators mentioned twice in the literature to list them all because it would make the table too long and difficult to read.

Only one of the included articles did not list any barriers [49], whereas three articles did not identify any facilitators [17,40,48]. For the analyzed publications, the count of distinct barriers reported ranged from 0 to 40 [24] (mean 10.24, SD 8.87, median 8). Likewise, between 0 and 48 facilitators [24] were mentioned in the literature (mean 9.18, SD 9.33, median 6).

Comparison of Expert Discussions and Literature Analysis

The combination of the expert discussions (15 barriers) and literature analysis (n=76) yielded a total of 77 specific barriers. In sum, 292 facilitators or success factors were found during the expert discussions (n=31) and via the literature analysis (n=268).

All barriers identified in the literature could be matched to the main categories (individual barriers, environmental barriers, technical barriers) defined in the Griebel et al study [26]. The technical barriers category was also mentioned in Mileski et al [24]. In mapping barriers from the expert discussions and literature analysis, we found that all but one barrier resulting from the expert discussions—system language (ie, the language of the service in use, such as German or English)—were also covered by the literature.

Facilitators derived from the literature could also be mapped completely to the adapted main categories from Griebel et al [26].

Discussion

Principal Results

Several references in the literature report specific barriers of eHealth adoption and implementation (eg, missing eHealth strategies [19,33,38]). However, distinct success factors are only included in one or two references (eg, clear governance of national eHealth strategy [32]).

Although 24 success factors from the expert discussions were not included in the results of our literature analysis, only one barrier identified in the expert discussions (system language) was not found via the literature analysis in this study. Thus, the overlap of success factors between the literature analysis and expert discussions was smaller than for barriers.

The top 10 barriers (see Table 1) and top six facilitators (see Table 2) as identified by their mention in the literature analysis can be named important factors influencing the implementation and adoption of eHealth services. The remaining factors seem to be more specific to certain stakeholders or areas of application since eHealth is a large field.

We identified many more success factors than barriers for the adoption of eHealth services. One reason behind this finding might be that success factors are outlined to a greater extent and in higher detail compared with barriers, which are reported very coarse-grained. Publication bias could be another reason. Unsuccessful projects tended not to analyze the reasons for failure, or at least not to publish their insights, compared with successful projects. For example, the Good eHealth Report [56] lists lessons learned, but the case studies published are only successful ones. For example, the reasons for the delay of the German Telematics Infrastructure and services are not published in scientific studies at all.

Limitations

The expert workshops were held only in Europe, which might have led to an underrepresentation of American, African, and Asian input to the discussions. Apart from PubMed, no further literature databases were consulted for this study. In addition, only one search with several parameters was conducted. However, the search parameters were adjusted several times to allow for more relevant articles to be found. Therefore, articles were randomly checked in the process. The literature analysis was restricted by search parameters including “barriers”’ and “facilitators” as well as “adoption” and “implementation,” which resulted in fewer articles found in the initial search. This also led to exclusion of fewer articles from the resulting list of publications because of irrelevance according to the chosen inclusion and exclusion criteria. Moreover, no white papers or reports by governments or other organizations were considered. Blog posts and articles by security professionals or operators and developers of these services, for example, were not included. A more comprehensive investigation with a focus on the aforementioned roles could consider these sources as well.

The categorization of barriers and facilitators was done by only one of the authors based on Griebel et al [26]. Thus, interrater reliability cannot be presented.

Comparison With Prior Work

The literature analysis included several systematic reviews conducted by other researchers. However, these reviews were

either focused on a specific eHealth application, disease, or patient subgroup. Kruse et al [20], for example, reviewed articles limited to military veterans with posttraumatic stress disorder to find out about factors that would influence telemedicine adoption. Mileski et al [24] focused their review on telemedicine for the self-management of hypertension. The systematic review by Ross et al [57] was limited to systematic reviews on factors influencing the implementation of eHealth published between 2009 and 2014. Ross et al searched with "MEDLINE, EMBASE, CINAHL, PsychINFO, and the Cochrane Library"—different databases than this study—so the relevant systematic reviews included in their study are only to a limited extent part of our literature analysis because we only searched PubMed. Ross et al used different categories for factors influencing the implementation of eHealth services informed by the Consolidated Framework for Implementation Research: innovation characteristics, outer setting, inner setting, characteristics of individuals, and process. Within these categories, Ross et al included components from all our top-level categories (individual, environmental and organizational, technical), such as adaptability and complexity (technical), and cost (environmental and organizational). Outer and inner setting, as described by Ross et al, would be included in environmental and organizational in our classification. However, Ross et al found "access to knowledge and information" to be a component of inner setting, which was added as an individual barrier (limited exposure/knowledge of eHealth) in our analysis. Another systematic review by O'Connor et al [43] analyzed qualitative studies to understand the factors affecting engagement with and recruitment to the use of eHealth applications. Bush et al [30] limited their systematic review to the pediatric population and the application type patient portal. The adoption of mHealth by health care professionals was the topic of the systematic review of Gagnon et al [33]. A narrative meta-review on e-mental health services was done by Batterham et al [17]. De Lusignan et al [32] did a literature review including electronic health records and patient access to health information, although eHealth applications were narrowed down to a subgroup.

In contrast to the studies included in our literature analysis, which were either based on literature analysis or reviews or experiences, we combined both expert discussions (experts' experience) and literature analysis. However, 24 success factors and one barrier from the expert discussions were not found in the literature. Also, the discussion groups "policy makers and politicians" and "data privacy officers and CIOs" could not be held due to a lack of participants.

Further approaches analyzed the applicability of the Technology Acceptance Model [13] and Unified Theory of Acceptance and Use of Technology [14,15] for the evaluation of eHealth services. However, these studies focused on contributing models for the evaluation of either eHealth services in general or a specific eHealth service instead of trying to provide a complete list of factors influencing their adoption. Models reflect only on certain details; they do not provide a holistic view of the impact factors for eHealth services.

Prior work includes analyses limited to within Europe, such as the Good eHealth Report [56] and MethoTelemed project [58]. The success factors given by the Good eHealth Report [56] are covered in the results of our literature analysis and expert discussions. Black et al [59] indicate that realizing the benefits of eHealth for quality and safety of health care is not guaranteed. They propose that more evaluation is necessary to identify all factors influencing eHealth services. The MethoTelemed project aimed to contribute to the evidence base on the impacts, benefits, and costs concerning telemedicine [58]. However, the project was constrained to telemedicine and focused mostly on methodological improvements.

In summary, the literature analysis conducted for this study, combined with findings from previous expert discussions, led to a more comprehensive list of barriers and facilitators for the adoption and implementation of eHealth services in general.

Conclusions

This work provides a comprehensive list of barriers and success factors based on two expert discussions and a literature analysis (see Multimedia Appendix 2). This list allows different stakeholders to address barriers and make use of facilitators in the planning phase of eHealth services. Thus, our work provides a valuable resource for health professionals, researchers, health care institutions, or consumers. With this resource, these groups might create better-suited applications and thus raise the adoption levels of consumer-centric eHealth services. Further studies on missing publications regarding the number of unsuccessful projects and eHealth services are necessary to research publication bias in this field.

Acknowledgments

We acknowledge financial support by Land Schleswig-Holstein within the funding program Open Access Publikationsfonds.

Authors' Contributions

BS did the literature analysis, contributed to the first expert discussion, participated in organizing the second expert discussion, combined the results from expert discussions and literature analysis, and wrote the paper draft. MP and MW organized the first and second expert discussion, gave input for the literature analysis, helped in combining the results from expert discussions and literature analysis, and reviewed and approved the paper. VS participated in the organization of the second expert discussion, gave input for the literature analysis, helped in combining the results from expert discussions and literature analysis, and reviewed and approved the paper. JS provided input, assisted in writing, and reviewed and approved the draft of the paper. BB gave input on the draft of the paper, and reviewed and approved the paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
List of all 56 analyzed publications including extracted barriers and facilitators.
[XLSX File (Microsoft Excel File), 31 KB - jmir_v21i11e14197_app1.xlsx]

Multimedia Appendix 2
Mindmap of all barriers and facilitators identified in expert discussions and literature.
[PDF File (Adobe PDF File), 3368 KB - jmir_v21i11e14197_app2.pdf]

References

1. Mitchell J. From Telehealth to e-Health: The Unstoppable Rise of e-Health. Canberra, Australia: Department of Communications, Information Technology and the Arts; 1999:55-56.
2. Wildenbos GA, Peute LW, Jaspers MW. Impact of patient-centered eHealth applications on patient outcomes: a review on the mediating influence of human factor issues. Yearb Med Inform 2016 Nov 10(1):113-119 [FREE Full text] [doi: 10.15265/IY-2016-031] [Medline: 27830238]
3. Catwell L, Sheikh A. Evaluating eHealth interventions: the need for continuous systemic evaluation. PLoS Med 2009 Aug;6(8):e1000126 [FREE Full text] [doi: 10.1371/journal.pmed.1000126] [Medline: 19688038]
4. Elbert NJ, van Os-Medendorp H, van Renselaar W, Ekeland AG, Hakkaart-van Roijen L, Raat H, et al. Effectiveness and cost-effectiveness of ehealth interventions in somatic diseases: a systematic review of systematic reviews and meta-analyses. J Med Internet Res 2014 Apr 16;16(4):e110 [FREE Full text] [doi: 10.2196/jmir.2790] [Medline: 24739471]
5. Choi WS, Park J, Choi JY, Yang J. Stakeholders' resistance to telemedicine with focus on physicians: utilizing the Delphi technique. J Telemed Telecare 2019 Jul 01;25(6):378-385. [doi: 10.1177/1357633X18775853] [Medline: 29792080]
6. Adler-Milstein J, Kvedar J, Bates DW. Telehealth among US hospitals: several factors, including state reimbursement and licensure policies, influence adoption. Health Aff (Millwood) 2014 Feb;33(2):207-215. [doi: 10.1377/hlthaff.2013.1054] [Medline: 24493762]
7. Colditz GA. Economic costs of obesity and inactivity. Med Sci Sports Exerc 1999 Nov;31(11 Suppl):S663-S667. [Medline: 10593542]
8. Crouter SE, Schneider PL, Karabulut M, Bassett DRJ. Validity of 10 electronic pedometers for measuring steps, distance, and energy cost. Med Sci Sports Exerc 2003 Aug;35(8):1455-1460. [doi: 10.1249/01.MSS.0000078932.61440.A2] [Medline: 12900704]
9. OECD. Focus on Spending on Health: Latest Trends. 2018 Jun. URL: http://www.oecd.org/health/health-systems/Health-Spending-Latest-Trends-Brief.pdf [accessed 2019-10-10]
10. Comstock J. Healthcare IT News. 2018 Jan 28. Apple to launch Health Records app with HL7's FHIR specifications at 12 hospitals. URL: https://www.healthcareitnews.com/news/apple-launch-health-records-app-hl7s-fhir-specifications-12-hospitals [accessed 2019-02-13] [WebCite Cache ID 76A8WdCVQ]
11. Naumann L, Babitsch B, Hübner U. Experts' insights into eHealth-legislation: comparing Switzerland and Germany. Stud Health Technol Inform 2018;253:148-152. [Medline: 30147061]
12. Davis FD, Bagozzi RP, Warshaw PR. User acceptance of computer technology: a comparison of two theoretical models. Manage Sci 1989 Aug;35(8):982-1003. [doi: 10.1287/mnsc.35.8.982]
13. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Quart 2003;27(3):425-478. [doi: 10.2307/30036540]
14. Griebel L, Sedlmayr B, Prokosch HU, Criegee-Rieck M, Sedlmayr M. Key factors for a successful implementation of personalized e-health services. Stud Health Technol Inform 2013;192:965. [Medline: 23920739]
15. Hoque R, Sorwar G. Understanding factors influencing the adoption of mHealth by the elderly: an extension of the UTAUT model. Int J Med Inform 2017 May;101:75-84. [doi: 10.1016/j.ijmedinf.2017.02.002] [Medline: 28347450]
16. Rho MJ, Choi IY, Lee J. Predictive factors of telemedicine service acceptance and behavioral intention of physicians. Int J Med Inform 2014 Aug;83(8):559-571. [doi: 10.1016/j.ijmedinf.2014.05.005] [Medline: 24961820]
17. Batterham PJ, Sunderland M, Calear AL, Davey CG, Christensen H, Teesson M, et al. Developing a roadmap for the translation of e-mental health services for depression. Aust N Z J Psychiatry 2015 Sep 23;49(9):776-784. [doi: 10.1177/0004867415582054] [Medline: 25907269]
18. Hadjistavropoulos HD, Nugent MM, Dirkse D, Pugh N. Implementation of internet-delivered cognitive behavior therapy within community mental health clinics: a process evaluation using the consolidated framework for implementation research. BMC Psychiatry 2017 Sep 12;17(1):331 [FREE Full text] [doi: 10.1186/s12888-017-1496-7] [Medline: 28899365]
19. Koivunen M, Hätönen H, Välimäki M. Barriers and facilitators influencing the implementation of an interactive Internet-portal application for patient education in psychiatric hospitals. Patient Educ Couns 2008 Mar;70(3):412-419. [doi: 10.1016/j.pec.2007.11.002] [Medline: 18079085]

20. Kruse CS, Atkins JM, Baker TD, Gonzales EN, Paul JL, Brooks M. Factors influencing the adoption of telemedicine for treatment of military veterans with post-traumatic stress disorder. J Rehabil Med 2018 May 08;50(5):385-392 [FREE Full text] [doi: 10.2340/16501977-2302] [Medline: 29700551]
21. Miake-Lye IM, Amulis A, Saliba D, Shekelle PG, Volkman LK, Ganz DA. Formative evaluation of the telecare fall prevention project for older veterans. BMC Health Serv Res 2011 May 23;11:119 [FREE Full text] [doi: 10.1186/1472-6963-11-119] [Medline: 21605438]
22. Mishuris RG, Stewart M, Fix GM, Marcello T, McInnes DK, Hogan TP, et al. Barriers to patient portal access among veterans receiving home-based primary care: a qualitative study. Health Expect 2015 Dec 12;18(6):2296-2305 [FREE Full text] [doi: 10.1111/hex.12199] [Medline: 24816246]
23. Band R, Bradbury K, Morton K, May C, Michie S, Mair FS, et al. Intervention planning for a digital intervention for self-management of hypertension: a theory-, evidence- and person-based approach. Implement Sci 2017 Feb 23;12(1):25 [FREE Full text] [doi: 10.1186/s13012-017-0553-4] [Medline: 28231840]
24. Mileski M, Kruse CS, Catalani J, Haderer T. Adopting telemedicine for the self-management of hypertension: systematic review. JMIR Med Inform 2017 Oct 24;5(4):e41 [FREE Full text] [doi: 10.2196/medinform.6603] [Medline: 29066424]
25. Shaw RJ, Kaufman MA, Bosworth HB, Weiner BJ, Zullig LL, Lee SD, et al. Organizational factors associated with readiness to implement and translate a primary care based telemedicine behavioral program to improve blood pressure control: the HTN-IMPROVE study. Implement Sci 2013 Sep 08;8:106 [FREE Full text] [doi: 10.1186/1748-5908-8-106] [Medline: 24010683]
26. Griebel L, Pobiruchin M, Wiesner M. Report on the MIE 2015 workshop: consumer health informatics-barriers and facilitators of eHealth usage among consumers. 2015 Aug Presented at: Report from: MIE2015 26th European Medical Informatics Conference; May 27-29, 2015; Madrid; p. 1-4. [doi: 10.13140/RG.2.1.2817.0081]
27. Ancker JS, Miller MC, Patel V, Kaushal R, HITEC Investigators. Sociotechnical challenges to developing technologies for patient access to health information exchange data. J Am Med Inform Assoc 2014;21(4):664-670 [FREE Full text] [doi: 10.1136/amiajnl-2013-002073] [Medline: 24064443]
28. Ariens LF, Schussler-Raymakers FM, Frima C, Flinterman A, Hamminga E, Arents BW, et al. Barriers and facilitators to eHealth use in daily practice: perspectives of patients and professionals in dermatology. J Med Internet Res 2017 Sep 05;19(9):e300 [FREE Full text] [doi: 10.2196/jmir.7512] [Medline: 28874336]
29. Bello AK, Molzahn AE, Girard LP, Osman MA, Okpechi IG, Glassford J, et al. Patient and provider perspectives on the design and implementation of an electronic consultation system for kidney care delivery in Canada: a focus group study. BMJ Open 2017 Mar 02;7(3):e014784 [FREE Full text] [doi: 10.1136/bmjopen-2016-014784] [Medline: 28255097]
30. Bush RA, Connelly CD, Fuller M, Pérez A. Implementation of the integrated electronic patient portal in the pediatric population: a systematic review. Telemed J E Health 2016 Mar 10;22(2):144-152 [FREE Full text] [doi: 10.1089/tmj.2015.0033] [Medline: 26258289]
31. Cook EJ, Randhawa G, Sharp C, Ali N, Guppy A, Barton G, et al. Exploring the factors that influence the decision to adopt and engage with an integrated assistive telehealth and telecare service in Cambridgeshire, UK: a nested qualitative study of patient 'users' and 'non-users'. BMC Health Serv Res 2016 Apr 19;16:137 [FREE Full text] [doi: 10.1186/s12913-016-1379-5] [Medline: 27095102]
32. de Lusignan S, Ross P, Shifrin M, Hercigonja-Szekeres M, Seroussi B. A comparison of approaches to providing patients access to summary care records across old and new europe: an exploration of facilitators and barriers to implementation. Stud Health Technol Inform 2013;192:397-401. [Medline: 23920584]
33. Gagnon MP, Ngangue P, Payne-Gagnon J, Desmartis M. m-Health adoption by healthcare professionals: a systematic review. J Am Med Inform Assoc 2016 Jan;23(1):212-220. [doi: 10.1093/jamia/ocv052] [Medline: 26078410]
34. Garg SK, Lyles CR, Ackerman S, Handley MA, Schillinger D, Gourley G, et al. Qualitative analysis of programmatic initiatives to text patients with mobile devices in resource-limited health systems. BMC Med Inform Decis Mak 2016 Feb 06;16:16 [FREE Full text] [doi: 10.1186/s12911-016-0258-7] [Medline: 26851941]
35. Hao WR, Hsu YH, Chen KC, Li HC, Iqbal U, Nguyen PA, et al. LabPush: a pilot study of providing remote clinics with laboratory results via short message service (SMS) in Swaziland, Africa - a qualitative study. Comput Methods Programs Biomed 2015 Jan;118(1):77-83. [doi: 10.1016/j.cmpb.2014.10.005] [Medline: 25453385]
36. Hunting G, Shahid N, Sahakyan Y, Fan I, Moneypenny CR, Stanimirovic A, et al. A multi-level qualitative analysis of Telehomecare in Ontario: challenges and opportunities. BMC Health Serv Res 2015 Dec 09;15:544 [FREE Full text] [doi: 10.1186/s12913-015-1196-2] [Medline: 26645639]
37. Jimbo M, Shultz CG, Nease DE, Fetters MD, Power D, Ruffin MT. Perceived barriers and facilitators of using a Web-based interactive decision aid for colorectal cancer screening in community practice settings: findings from focus groups with primary care clinicians and medical office staff. J Med Internet Res 2013 Dec 18;15(12):e286 [FREE Full text] [doi: 10.2196/jmir.2914] [Medline: 24351420]
38. Lennon MR, Bouamrane MM, Devlin AM, O'Connor S, O'Donnell C, Chetty U, et al. Readiness for delivering digital health at scale: lessons from a longitudinal qualitative evaluation of a national digital health innovation program in the United Kingdom. J Med Internet Res 2017 Feb 16;19(2):e42 [FREE Full text] [doi: 10.2196/jmir.6900] [Medline: 28209558]

39. Levine M, Richardson JE, Granieri E, Reid MC. Novel telemedicine technologies in geriatric chronic non-cancer pain: primary care providers' perspectives. Pain Med 2014 Feb;15(2):206-213 [FREE Full text] [doi: 10.1111/pme.12323] [Medline: 24341423]
40. Luque AE, van Keken A, Winters P, Keefer MC, Sanders M, Fiscella K. Barriers and facilitators of online patient portals to personal health records among persons living with HIV: formative research. JMIR Res Protoc 2013 Jan 22;2(1):e8 [FREE Full text] [doi: 10.2196/resprot.2302] [Medline: 23612564]
41. Makai P, Perry M, Robben SH, Schers H, Heinen M, Olde Rikkert MG, et al. Which frail older patients use online health communities and why? A mixed methods process evaluation of use of the Health and Welfare portal. J Med Internet Res 2014 Dec 17;16(12):e278 [FREE Full text] [doi: 10.2196/jmir.3609] [Medline: 25519769]
42. Molfenter T, Boyle M, Holloway D, Zwick J. Trends in telemedicine use in addiction treatment. Addict Sci Clin Pract 2015 May 28;10:14 [FREE Full text] [doi: 10.1186/s13722-015-0035-4] [Medline: 26016484]
43. O'Connor S, Hanlon P, O'Donnell CA, Garcia S, Glanville J, Mair FS. Understanding factors affecting patient and public engagement and recruitment to digital health interventions: a systematic review of qualitative studies. BMC Med Inform Decis Mak 2016 Sep 15;16(1):120 [FREE Full text] [doi: 10.1186/s12911-016-0359-3] [Medline: 27630020]
44. Ochoa AIII, Kitayama K, Uijtdehaage S, Vermillion M, Eaton M, Carpio F, et al. Patient and provider perspectives on the potential value and use of a bilingual online patient portal in a Spanish-speaking safety-net population. J Am Med Inform Assoc 2017 Nov 01;24(6):1160-1164. [doi: 10.1093/jamia/ocx040] [Medline: 28460130]
45. Odnoletkova I, Buysse H, Nobels F, Goderis G, Aertgeerts B, Annemans L, et al. Patient and provider acceptance of telecoaching in type 2 diabetes: a mixed-method study embedded in a randomised clinical trial. BMC Med Inform Decis Mak 2016 Nov 09;16(1):142 [FREE Full text] [doi: 10.1186/s12911-016-0383-3] [Medline: 27825340]
46. Popeski N, McKeen C, Khokhar B, Edwards A, Ghali WA, Sargious P, et al. Perceived barriers to and facilitators of patient-to-provider e-mail in the management of diabetes care. Can J Diabetes 2015 Dec;39(6):478-483. [doi: 10.1016/j.jcjd.2015.07.001] [Medline: 26409770]
47. Radhakrishnan K, Xie B, Jacelon CS. Unsustainable home telehealth: a Texas qualitative study. Gerontologist 2016 Oct;56(5):830-840. [doi: 10.1093/geront/gnv050] [Medline: 26035878]
48. Sarfo FS, Adamu S, Awuah D, Sarfo-Kantanka O, Ovbiagele B. Potential role of tele-rehabilitation to address barriers to implementation of physical therapy among West African stroke survivors: A cross-sectional survey. J Neurol Sci 2017 Oct 15;381:203-208 [FREE Full text] [doi: 10.1016/j.jns.2017.08.3265] [Medline: 28991682]
49. Schickedanz A, Huang D, Lopez A, Cheung E, Lyles CR, Bodenheimer T, et al. Access, interest, and attitudes toward electronic communication for health care among patients in the medical safety net. J Gen Intern Med 2013 Jul;28(7):914-920 [FREE Full text] [doi: 10.1007/s11606-012-2329-5] [Medline: 23423453]
50. Swinkels IC, Huygens MW, Schoenmakers TM, Oude Nijeweme-D'Hollosy W, van Velsen L, Vermeulen J, et al. Lessons learned from a living lab on the broad adoption of eHealth in primary health care. J Med Internet Res 2018 Mar 29;20(3):e83 [FREE Full text] [doi: 10.2196/jmir.9110] [Medline: 29599108]
51. Tieu L, Sarkar U, Schillinger D, Ralston JD, Ratanawongsa N, Pasick R, et al. Barriers and facilitators to online portal use among patients and caregivers in a safety net health care system: a qualitative study. J Med Internet Res 2015;17(12):e275 [FREE Full text] [doi: 10.2196/jmir.4847] [Medline: 26681155]
52. van Cranenburgh OD, Ter Stege JA, de Korte J, de Rie MA, Sprangers MA, Smets EM. Patient-reported outcome measurement in clinical dermatological practice: relevance and feasibility of a web-based portal. Dermatology 2016;232(1):64-70 [FREE Full text] [doi: 10.1159/000440613] [Medline: 26513574]
53. van der Meij E, Huirne JA, Ten Cate AD, Stockmann HB, Scholten PC, Davids PH, et al. A perioperative eHealth program to enhance postoperative recovery after abdominal surgery: process evaluation of a randomized controlled trial. J Med Internet Res 2018 Jan 02;20(1):e1 [FREE Full text] [doi: 10.2196/jmir.8338] [Medline: 29295808]
54. Varsi C, Ekstedt M, Gammon D, Ruland CM. Using the consolidated framework for implementation research to identify barriers and facilitators for the implementation of an internet-based patient-provider communication service in five settings: a qualitative study. J Med Internet Res 2015 Nov 18;17(11):e262 [FREE Full text] [doi: 10.2196/jmir.5091] [Medline: 26582138]
55. Wildenbos GA, Peute L, Jaspers M. Facilitators and barriers of electronic health record patient portal adoption by older adults: a literature study. Stud Health Technol Inform 2017;235:308-312. [Medline: 28423804]
56. European Commission. eHealth in Action-Good Practice in European Countries. Good eHealth Report. Luxembourg: Office for Official Publications of the European Communities; 2009.
57. Ross J, Stevenson F, Lau R, Murray E. Factors that influence the implementation of e-health: a systematic review of systematic reviews (an update). Implement Sci 2016 Oct 26;11(1):146 [FREE Full text] [doi: 10.1186/s13012-016-0510-7] [Medline: 27782832]
58. Ekeland AG, Bowes A, Flottorp S. Methodologies for assessing telemedicine: a systematic review of reviews. Int J Med Inform 2012 Jan;81(1):1-11. [doi: 10.1016/j.ijmedinf.2011.10.009] [Medline: 22104370]
59. Black AD, Car J, Pagliari C, Anandan C, Cresswell K, Bokun T, et al. The impact of eHealth on the quality and safety of health care: a systematic overview. PLoS Med 2011 Jan 18;8(1):e1000387 [FREE Full text] [doi: 10.1371/journal.pmed.1000387] [Medline: 21267058]

Abbreviations

CIO: chief information officer
eHID: eHealth Innovation Days Conference
MIE: Medical Informatics Europe Conference

Edited by CL Parra-Calderón; submitted 29.03.19; peer-reviewed by MVR González, E Brainin, E Neter; comments to author 23.05.19; revised version received 09.07.19; accepted 01.09.19; published 22.11.19.

Please cite as:
Schreiweis B, Pobiruchin M, Strotbaum V, Suleder J, Wiesner M, Bergh B
Barriers and Facilitators to the Implementation of eHealth Services: Systematic Literature Analysis
J Med Internet Res 2019;21(11):e14197
URL: http://www.jmir.org/2019/11/e14197/
doi:10.2196/14197
PMID:31755869

Original Paper

User Engagement and Attrition in an App-Based Physical Activity Intervention: Secondary Analysis of a Randomized Controlled Trial

Sarah Edney[1], BA; Jillian C Ryan[2], PhD; Tim Olds[1], PhD; Courtney Monroe[3], PhD; François Fraysse[1], PhD; Corneel Vandelanotte[4], PhD; Ronald Plotnikoff[5], PhD; Rachel Curtis[1], PhD; Carol Maher[1], PhD

[1]Alliance for Research in Exercise, Nutrition and Activity, University of South Australia, Adelaide, Australia

[2]Health and Biosecurity, Commonwealth Scientific and Industrial Research Organisation, Adelaide, Australia

[3]Arnold School of Public Health, University of South Carolina, Columbia, SC, United States

[4]Physical Activity Research Group, Appleton Institute, School of Health, Medical and Applied Sciences, Central Queensland University, Rockhampton, Australia

[5]Priority Research Centre for Physical Activity and Nutrition, The University of Newcastle, Newcastle, Australia

Corresponding Author:
Sarah Edney, BA
Alliance for Research in Exercise, Nutrition and Activity
University of South Australia
North Terrace
Adelaide
Australia
Phone: 61 883026611
Email: Sarah.Edney@mymail.unisa.edu.au

Abstract

Background: The success of a mobile phone app in changing health behavior is thought to be contingent on engagement, commonly operationalized as frequency of use.

Objective: This subgroup analysis of the 2 intervention arms from a 3-group randomized controlled trial aimed to examine user engagement with a 100-day physical activity intervention delivered via an app. Rates of engagement, associations between user characteristics and engagement, and whether engagement was related to intervention efficacy were examined.

Methods: Engagement was captured in a real-time log of interactions by users randomized to either a gamified (n=141) or nongamified version of the same app (n=160). Physical activity was assessed via accelerometry and self-report at baseline and 3-month follow-up. Survival analysis was used to assess time to nonuse attrition. Mixed models examined associations between user characteristics and engagement (total app use). Characteristics of super users (top quartile of users) and regular users (lowest 3 quartiles) were compared using *t* tests and a chi-square analysis. Linear mixed models were used to assess whether being a super user was related to change in physical activity over time.

Results: Engagement was high. Attrition (30 days of nonuse) occurred in 32% and 39% of the gamified and basic groups, respectively, with no significant between-group differences in time to attrition (P=.17). Users with a body mass index (BMI) in the healthy range had higher total app use (mean 230.5, 95% CI 190.6-270.5; F_2=8.67; P<.001), compared with users whose BMI was overweight or obese (mean 170.6, 95% CI 139.5-201.6; mean 132.9, 95% CI 104.8-161.0). Older users had higher total app use (mean 200.4, 95% CI 171.9-228.9; F_1=6.385; P=.01) than younger users (mean 155.6, 95% CI 128.5-182.6). Super users were 4.6 years older (t_{297}=3.6; P<.001) and less likely to have a BMI in the obese range (χ^2_2=15.1; P<.001). At the 3-month follow-up, super users were completing 28.2 (95% CI 9.4-46.9) more minutes of objectively measured physical activity than regular users ($F_{1,272}$=4.76; P=.03).

Conclusions: Total app use was high across the 100-day intervention period, and the inclusion of gamified features enhanced engagement. Participants who engaged the most saw significantly greater increases to their objectively measured physical activity over time, supporting the theory that intervention exposure is linked to efficacy. Further research is needed to determine whether these findings are replicated in other app-based interventions, including those experimentally evaluating engagement and those conducted in real-world settings.

Trial Registration: Australian New Zealand Clinical Trials Registry ACTRN12617000113358; https://www.anzctr.org.au/ACTRN12617000113358.aspx

(J Med Internet Res 2019;21(11):e14645) doi:10.2196/14645

KEYWORDS

physical activity; smartphone; behavior

Introduction

Background

Mobile phone apps have been proposed as a cost-effective method of delivering wide-scale, appealing interventions, targeting lifestyle-related health problems, such as physical inactivity, to prevent chronic diseases [1,2]. However, despite the rapid growth of these approaches, researchers only started to develop and examine custom-made apps to deliver physical activity (PA) interventions in the past decade [3]. To date, these have tended to report mixed results [4,5], with meta-analyses indicating overall modest effects [6,7]. Contributing to overall modest efficacy are the lower rates of intervention exposure and compliance and the high levels of attrition that are often reported [8].

The apparent link between intervention exposure and efficacy [8-10] highlights the need for detailed understanding of user engagement, which considers how frequently users access or use different features of the app or simply log on to the app at all. To date, engagement has been conceptualized and operationalized in different ways, and attempts have been made to reach a shared understanding of approaches [11,12]. Research studies have tended to focus on frequency of app use and dropout rates assessed as percentages of users who cease using the app during the intervention [12,13]. Others take a user-centric approach, considering the appeal of an app and the experience of using it reported directly from users themselves, via interviews or questionnaires [14]. Commercial companies, on the other hand, view app engagement differently, focusing on metrics that include the number of daily active users (DAUs), the number of monthly active users (MAUs), and *churn*, measured either as the rate at which nonuse attrition occurs or, to take into account new app downloads, by dividing the number of users at the end of a period by the number of users present at the beginning and expressing this as a percentage [15]. Finally, commercial companies examine the characteristics and usage patterns of *super users*, who use an app the most in terms of frequency, time spent, or who remain active users for an extended period. These super users are more likely to rate an app favorably, share it with their friends, and engage with most of its features [16-19]. More recently, as researchers look to disseminate apps with demonstrated efficacy into natural, ecologically valid settings [20-22], it may be more useful to adopt these commercial metrics. In this approach, engagement metrics used to measure the success of for-profit commercial apps may be more informative or provide unique insights and could be adopted to allow for comparison between research- and industry-led PA apps.

Objectives

This study therefore sought to contribute to addressing this need by applying engagement metrics from commercial settings to app usage data collected within a health behavior randomized controlled trial (RCT) to provide insights regarding usage rates and associations between user characteristics and app usage.

This study aimed to (1) describe user engagement with a PA app as total app use, DAU and MAU, and nonuse attrition (*churn rate*), (2) examine whether user characteristics were associated with engagement (total app use), (3) compare characteristics between *super users* and *regular users*, and (4) examine whether total app use or being a super user was related to intervention efficacy.

Methods

Research Design

This study is a secondary and subgroup analysis of the 2 intervention arms from a 3-group cluster RCT that evaluated the efficacy of an app-based intervention in increasing performance of moderate-to-vigorous PA (MVPA), in comparison with a waitlist control group. Full details of the RCT have been published previously [23]. The study received ethics approval from the University of South Australia's Human Research Ethics Committee (protocol: 33967), and it was registered with the Australian and New Zealand Clinical Trial Registry (protocol: 12617000113358).

Participants and Procedures

Recruitment took place in 2016 and 2017, primarily via a paid Facebook advertising campaign and free advertisements placed in community-based Facebook groups. Participants were eligible if they were aged 18 to 65 years, lived anywhere in Australia, used Facebook weekly, reported as completing less than 150 min of MVPA per week, and were able to sign up to the study in a group of at least 3 and a maximum of 8 existing Facebook friends who also met the study inclusion criteria and were willing to join the study.

Participant teams were randomized to either the gamified or basic app (descriptions below) or to the waitlist control group in a 1:1:1 allocation ratio. The gamified app included social and gamified features designed to encourage use of the app and in-app interaction within teams of friends using the app together, whereas the basic app was designed to be used by each participant independently. Participant teams allocated to the waitlist control group are not included in the current analysis.

Intervention

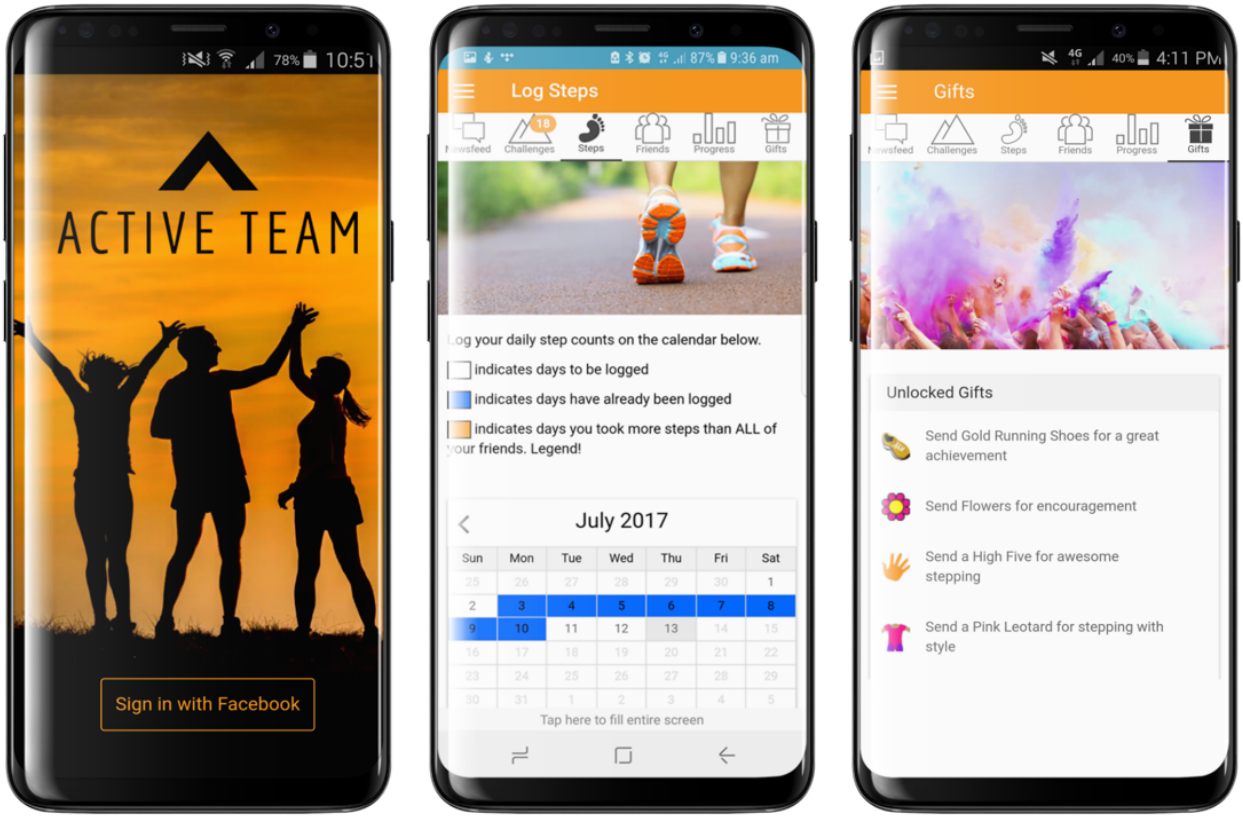
Users in both app groups received a pedometer (Zencro, TW64S) to measure their daily step counts**,** and they received

a weekly email with a summary of their individual step count progress. For gamified users, the email also highlighted a different app feature each week.

Gamified App

The gamified app (*Active Team*, see Figure 1) encouraged users to take 10,000 steps per day for 100 days. This app was designed by the research team (led by CM) and developed in conjunction with a software development company (Portal Australia) primarily for academic research purposes. Figure 1 shows screenshots of the app. Full app details and additional images are available in the study protocol [23]. Within the app, users could log their steps into a calendar that tracked their progress over time. Simple gamification (game-like elements included in nongame settings [24]) features were included: a leaderboard, unlockable gifts and medals, a Facebook-style newsfeed, and mini-PA challenges. The leaderboard displayed a ranked summary of each teammates' step count progress. New virtual gifts and medals were unlocked when users reached predetermined step counts. The Facebook-style newsfeed allowed users to share messages with each other, and mini challenges could be sent between users to encourage short bursts of PA (eg, to take 2000 steps in the next 20 min or to take 12,000 steps per day for 3 consecutive days). Users received a daily push notification reminding them to log their step counts, and users received push notifications when a teammate interacted with them within the app.

Figure 1. Screenshots of the Active Team app Showing (left to right): splashscreen, step calendar, and virtual gifts.



Basic App

Users of the basic app received a pared down version of the Active Team app that allowed them to enter and monitor their own steps in the calendar, and they received a daily push notification reminding them to enter their step counts. This version of the app contained no gamified or social features.

Outcomes

This study used demographic and PA data, as well as app usage data from the 100-day intervention period.

Demographics

Demographic characteristics were collected within a Web-based survey at baseline. Users reported their sex (male and female), age—median=41 and categorized into older (>41) or younger (<40)—the highest level of education they had achieved (high school or less, some college or further education institution, and university degree or higher), and height and weight, from which body mass index (BMI; kg/m^2) was calculated and converted into categories: healthy weight (<24.9), overweight (25-29.9), obese (>30).

Physical Activity

PA was assessed at baseline (before being randomized to the gamified or basic app) and 3-month follow-up (to coincide with the end of the 100-day intervention period), using GENEActiv accelerometer. At both time points, participants received an accelerometer, instructions to wear the device for 7 days, and a reply-paid envelope, via the Australian postal system. These data were classified into MVPA by using cut points established by Esliger et al [25]. Self-reported PA was collected via the Active Australia Survey (AAS) [26], delivered electronically.

The AAS collects details of minutes spent in walking and moderate and vigorous activity in the past week and calculates MVPA from moderate and vigorous activity.

Engagement

Engagement data were collected using a real-time log of each user's interactions with the app, which was automatically uploaded to the study's server over the 100-day intervention period. The operationalization of the engagement metrics used in this study is provided below.

Total App Use

The primary engagement metric used in this study is the number of times the app features were used (*total app use*) during the 100-day intervention period. To calculate this individual-level engagement metric, each interaction with a feature of the app (ie, step calendar, newsfeed, challenge page, gift page, and friends page) was included as a single use. Where the server recorded multiple interactions with a single app feature within a 15-min period, this was counted once only.

Daily Active Users

This sample-level engagement metric was calculated as the number and percentage of gamified and basic app users who accessed the app each day.

Monthly Active Users

This sample-level engagement metric was calculated as the number and percentage of gamified and basic app users who accessed the app at least once every 30 days.

Nonuse Attrition

Attrition was defined as occurring once the user ceased accessing the app for 30 consecutive days or more; the nonuse attrition rate is a sample-level engagement metric. The 30-day threshold is often used to describe the use of commercial apps [15,27]. A sensitivity analysis was performed using the threshold of 14 days of nonuse, a time frame, which has been previously applied in research studies [22,28,29].

Superusers

Super users were defined as users whose total app use fell into the top quartile of all users (individual-level engagement metric).

Step Calendar Use

The number of unique visits to the step calendar (*step calendar use*) was also calculated as an individual-level engagement metric, and this has been presented descriptively.

DAU, MAU, nonuse attrition, and super users of the step calendar were calculated, as users of both apps could access this feature, whereas users of the gamified app had access to additional gamified and social features, which could potentially increase their total app use (and were designed to achieve this). The results from the 2 metrics (total app use and step calendar use) were near identical; therefore, the results pertaining to step calendar use are provided in Multimedia Appendix 1.

Statistical Analysis

Demographic characteristics, baseline PA levels, total app use, DAU, and MAU are presented descriptively. Kaplan-Meier survival curves [30] were used to assess the time at which attrition (30 consecutive days of nonuse; sensitivity analysis: 14 consecutive days of nonuse) from each app occurred. The number of days of app use was the time variable, and the event variable was specified as being when the user ceased using the app for 30 or more consecutive days (or 14 or more consecutive days for the sensitivity analysis). Observations were classified as censored when the app was still being used by the end of the 100-day intervention period. The log-rank test was used to determine whether the time until nonuse attrition occurred was statistically significantly different between groups (gamified or basic app group).

Associations between user characteristics (sex, education, BMI, age, and group [gamified or basic app] and engagement [total app use]) were examined using the general linear model procedure. Significant interactions for categorical variables (education and BMI) were followed up with post hoc pairwise comparisons (Kruskal-Wallis tests with Bonferroni correction), where appropriate. Demographic characteristics and patterns of app usage by super users (top quartile of users) and regular users (lowest 3 quartiles) are presented descriptively. Between-group differences were analyzed using independent *t* tests for continuous variables (age, objective, and self-reported PA) and chi-square tests for categorical and binary variables (education, BMI category, and sex). Linear mixed models were used to examine whether total app use or being a super user was related to changes in objective or self-reported PA over time. Total app use, time, and a total app use-by-time interaction were entered as fixed effects in the first model, and superuser status (ie, yes or no), time, and a superuser status-by-time interaction were entered as the fixed effects in the second model. In both models, individual and team were entered as random effects. Adjustments were made for demographic characteristics (sex, education, BMI, and age) by entering these variables into the model. There were no substantial differences between the adjusted and unadjusted models, and the unadjusted models are presented. All analyses were undertaken in SPSS version 25 (IBM Corp).

Results

User Characteristics

A total of 301 users were randomized to either the gamified (n=141) or basic (n=160) version of the app. Users in both groups were predominantly female (73.8%, 222/301), university educated (53.2%, 159/301), overweight or obese (35.8%, 107/301 and 43.1%, 129/301), and with an average age of 42 (SD 12) years. There were no statistically significant differences in baseline characteristics between the 2 groups. Overall use of both apps was high. On an average, total app use for the gamified group was 239 (SD 202) compared with 120 (SD 94) for basic app users. Users accessed the step calendar on an average of 44 (SD 32) and 36 (SD 29) times for the gamified and basic apps, respectively. Full details of user characteristics and app usage are presented in Table 1.

Table 1. Participant baseline characteristics, total app use, and total step calendar use.

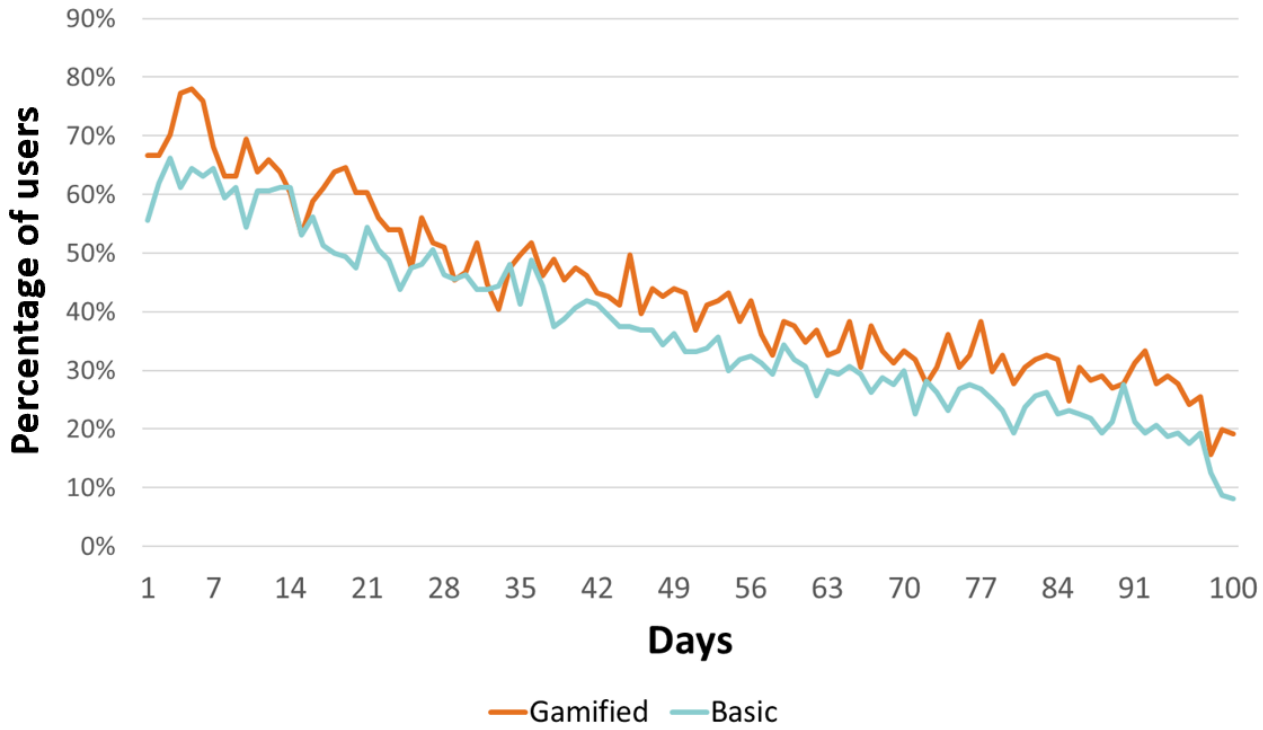
Participant characteristic	Gamified (n=141)	Basic (n=160)	All (N=301)
Sex, n (%)			
Female	106 (75.2)	116 (72.5)	222 (73.8)
Education level, n (%)			
High school or less	26 (18.7)	25 (15.6)	51 (17.1)
Some college	36 (25.9)	53 (33.1)	89 (29.8)
University degree	77 (55.4)	82 (51.3)	159 (53.2)
Body mass index, n (%)			
Healthy weight	26 (18.7)	37 (23.1)	63 (21.1)
Overweight	52 (37.4)	55 (34.4)	107 (35.8)
Obese	61 (43.9)	68 (42.5)	129 (43.1)
Age (years), mean (SD)	43.3 (11.5)	40.4 (12.2)	41.8 (11.9)
Objective MVPA[a] minute per day, mean (SD)	106.7 (52.9)	102.7 (50.6)	104.6 (51.6)
Self-reported MVPA minute per week, mean (SD)	243.6 (211.7)	262.1 (264.2)	253.5 (240.9)
Total app use, mean (SD)	239.0 (202.0)	120.17 (94.3)	175.3 (165.2)
Step calendar use, mean (SD)	43.6 (32.4)	36.1 (28.5)	39.5 (30.6)

[a]MVPA: moderate-to-vigorous physical activity.

Daily Active Users

Figure 2 shows the sample-level engagement metric of number of users accessing the app each day. The number of DAUs declined over the 100-day intervention period. At day 7, 68.0% (96/141) of the gamified and 64.4% (102/160) of the basic group were accessing the app daily. By day 49, this declined to 43.9% (62/141) of the gamified and 36.2% (58/160) of the basic group. By day 91, 31.2% (44/141) of the gamified and 21.2% (34/160) of the basic group were accessing the app daily. Differences between groups were statistically significant (t=14.96, *P*<.001), favoring the gamified group.

Figure 2. Daily active users and total app use.



Monthly Active Users

The sample-level engagement metric, number of MAUs, remained high across the 3-month intervention period. In the first 30 days, 93.7% (282/301) of all users accessed the app (94.3%, 133/141) and 93.1% (149/160) of the gamified and basic groups, respectively). This number declined slightly across the intervention period, but it remained high at 76.4% (230/301) for both groups (75.9%, 107/141) and 76.9% (123/160) of the gamified and basic groups, respectively). Differences in MAUs between the 2 groups were not statistically significant (*P*=.99).

Nonuse Attrition

When assessed using the 30 days of nonuse threshold, nonuse attrition occurred for 31.9% (96/141) and 39.4% (97/160) of the gamified and basic groups, respectively, with no significant differences between groups in the time to attrition (log-rank test, *P*=.17; see Figure 3). Sensitivity analysis applied the 14 days of nonuse threshold and found that nonuse attrition occurred for 48.9% (69/141) and 58.7% (94/160) of the gamified and basic groups, respectively, and differences were not statistically significant (log-rank test, *P*=.12; see Figure 4).

Figure 3. Kaplan-Meier survival estimates showing time to nonuse attrition, 30-day nonuse threshold, and total app use.

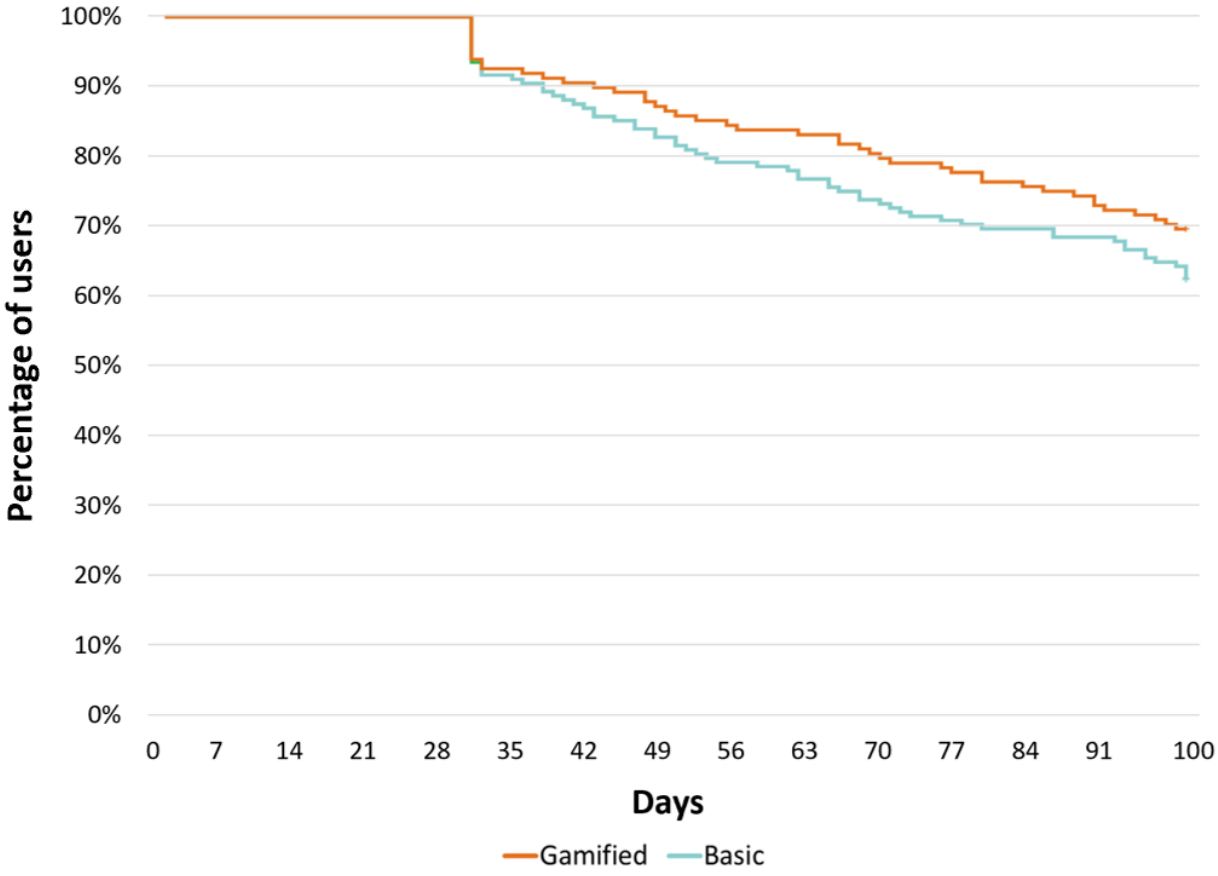
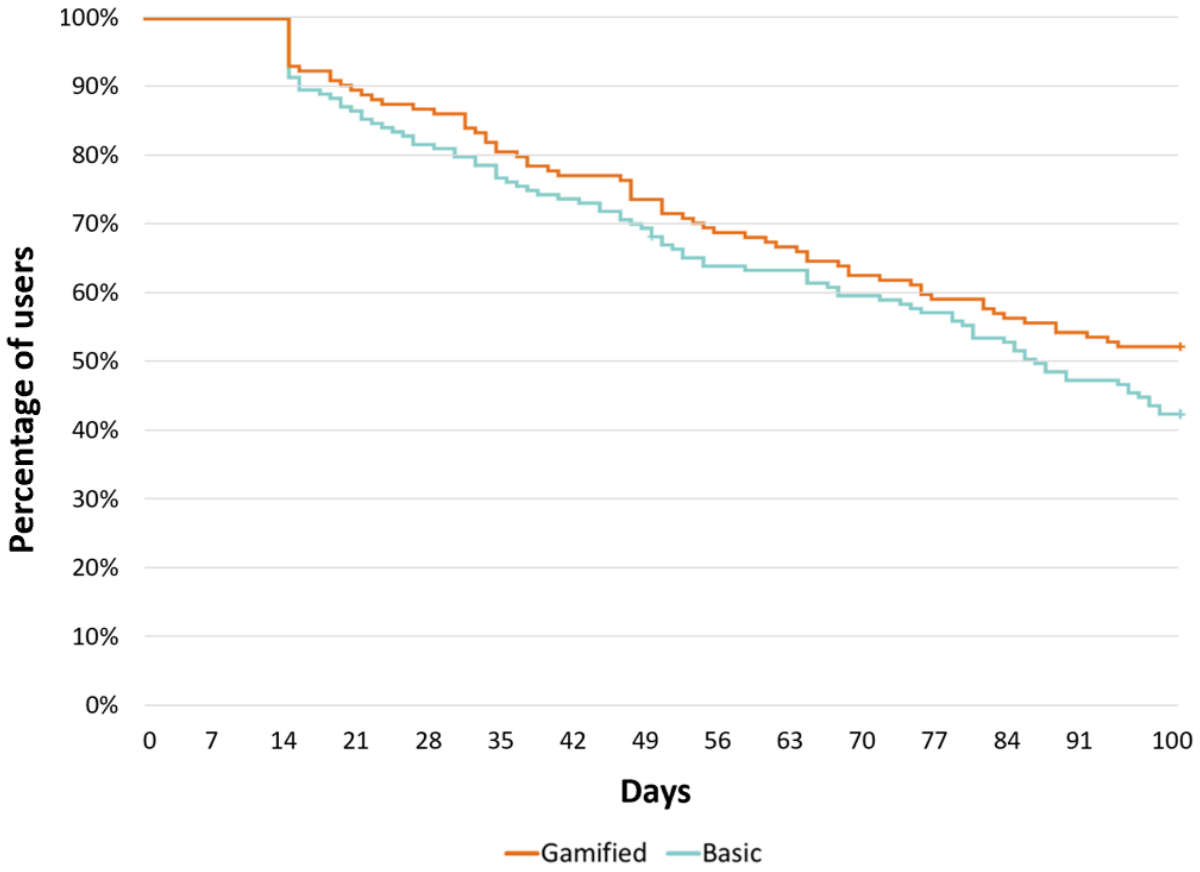


Figure 4. Kaplan-Meier survival estimates showing time to non-use attrition 14-day non-use threshold - total app use.



User Characteristics and Total App Use

Associations between user characteristics and total app use (individual-level engagement metric) were examined using general linear models. Results indicated that BMI, age, and app group (gamified or basic) were each associated with total app use. Post hoc pairwise comparison (Kruskal-Wallis test) indicated total app use varied on the basis of BMI (χ^2_2=12.8, *P*=.001), driven by differences between those with a BMI classified as healthy (median total app use=194), compared with obese (median total app use=112; *P*=.001). Older users had higher total app use (mean 200.4, 95% CI 171.9-228.9; F_1=6.39; *P*=.01) than younger users (mean=155.6, 95% CI 128.5-182.6). Users of the gamified app also had more total app use (mean=236.5, 95% CI 207.9-265.1; F_1=45.12; *P*<.001), when compared with those using the basic app (mean=119.5, 95% CI 93.0-146.0). Full details of all associations are presented in Table 2.

Table 2. Associations between user characteristics and total app use.

Characteristics	Total, n	Total app use		
		Mean (95% CI)	*F* test *(df)*	*P* value
Sex				
Male	78	163.7 (129.1-198.2)	2.10 *(1)*	.15
Female	221	192.3 (170.1-214.6)	—[a]	—
Education level				
High school or less	51	170.8 (128.7-212.9)	2.21 *(2)*	.11
Some college	89	162.0 (128.0-195.0)	—	—
University degree	159	201.2 (174.9-227.5)	—	—
Body mass index				
Healthy weight	63	230.5 (190.6-270.5)	8.67 (2)	*<.001* [b]
Overweight	107	170.6 (139.5-201.6)	—	—
Obese	129	132.9 (104.8-161.0)	—	—
Age (years)[c]				
Younger	149	155.6 (128.5-182.6)	6.39 *(1)*	*.01*
Older	150	200.4 (171.9-228.9)	—	—
Group				
Gamified	141	236.5 (207.9-265.1)	45.12 *(1)*	*<.001*
Basic	160	119.5 (93.0-146.0)	—	—

[a]Not applicable.

[b]*Italics* indicates statistical significance.

[c]Age: younger=≤40; older=≥41.

Super Versus Regular Users

Super users used the app an average of 401 (SD 160; range 248 to 1062) times compared with 100 (SD 71; range 0 to 239) times for regular users. Super users were more likely to be a gamified rather than basic app user (χ^2_1=29.4; *P*<.001), were 4.6 years older (t_{297}=3.6; *P*<.001), and were more likely to have a BMI that placed them within the healthy or overweight, rather than obese, range (χ^2_1=15.1; *P*<.001), when compared with regular users.

There were no significant differences between super and regular users for either objectively measured or self-reported PA data at baseline; however, super users had greater increases in their objectively measured PA change scores (t_{239}=4.1; *P*<.001), when compared with regular users whose PA decreased slightly from baseline. Full details of differences between super and regular users with regard to baseline characteristics and PA change over time can be found in Table 3.

Table 3. Baseline characteristic and physical activity change score comparisons between super and regular users.

Characteristics	Total app use		
	Super	Regular	*P* value
Sex, n (%)			
Male	15 (19.7)	64 (28.4)	.14
Female	61 (80.3)	161 (71.6)	—[a]
Education level, n (%)			
High school or less	9 (12.0)	42 (18.8)	.09
Some college	18 (24.0)	71 (31.7)	—
University degree	48 (64.0)	111 (49.6)	—
Body mass index, n (%)			
Healthy weight	22 (29.3)	41 (18.3)	*<.001* [b]
Overweight	35 (46.7)	72 (32.1)	—
Obese	18 (24.0)	111 (49.6)	—
Age (years), mean (SD)	46.0 (11.5)	40.4 (11.8)	*<.001*
Group, n (%)			
Gamified	56 (73.7)	85 (37.8)	*<.001*
Basic	20 (26.3)	140 (62.2)	—
Objective PA[c] minute per day at baseline, mean (SD)	107.0 (45.7)	103.7 (53.6)	.63
Self-reported PA minute per week at baseline, mean (SD)	244.0 (211.2)	256.7 (250.5)	.69
Objective PA minute per day change score, mean (SD)[d]	18.4 (45.0)	–8.1 (47.1)	*<.001*
Self-reported PA minute per week change score, mean (SD)[d]	239.7 (424.4)	157.3 (388.0)	.14

[a]Not applicable.

[b]*Italics* indicates statistical significance.

[c]PA: physical activity.

[d]Physical activity change scores from baseline to 3-month follow-up (end of intervention).

Total App Use and Changes to Physical Activity

There was a weak, significant total app use-by-time interaction effect for objective PA ($F_{1,272}$=4.5; P=.04) and self-reported PA ($F_{1,304}$=6.56; P=.01), where higher total app use (individual-level engagement metric) was associated with greater increases in PA at 3-month follow-up.

Super Users and Changes to Physical Activity

At the 3-month follow-up, objective MVPA had increased from baseline for super users of the app, whereas it decreased slightly for regular users. There was a significant group by time interaction, where super users were completing 28.2 (SE 9.5, 95% CI 9.4-46.9) more min of MVPA than regular users ($F_{1,272}$=4.76; P=.03). Differences between super and regular users for self-reported MVPA favored the super users (mean 89.7, SE 43.4, 95% CI 4.4-175.1); however, did not reach statistical significance ($F_{1,297}$=3.31; P=.07).

Discussion

Principlal Findings

The aim of this study was to examine user engagement with an app-based PA intervention. Use of the app was high, and although this trended downward, the low attrition rates indicate that most users were returning to the app at least once every 30 days throughout the intervention. Rates of use differed on the basis of demographic characteristics, where older users and those with a BMI in the healthy range used the app more. Unsurprisingly, engagement was higher for users of the gamified app, which included additional features designed to encourage this. Super users (top quartile of users) also tended to be older and were less likely to be obese. Super users increased their PA over time, whereas regular users decreased.

Engagement with the app was high compared with both research- and industry-led apps. In the gamified and basic app groups combined, users accessed the app features on an average of 175 (SD 165) times during the 100 days. In comparison, a systematic review reported rates of engagement ranging between 5 and 55 times (5%-15% of intended engagement) within Web-based health behavior interventions, ranging in duration

from 3 months to 26 weeks [8]. Although engagement data from commercial apps are not readily available, a study of over a million users of a commercial weight loss app (Lose It!) that operationalized engagement as number of days of use found mean engagement of 29 days, which increased to 172 days for users accessing customizable features in the app (eg, personalized PA plan) [31]. Other reports examining all apps available in the Apple and Google Play app stores suggest that 21% of all downloaded apps are engaged with just once in the first 6 months [32]. Declining rates of engagement over time are often reported by researcher-led Web-based health interventions [8,9,33,34]; however, although our daily engagement dropped from a high of 70.8% (213/301) to a low of 13.3% (40/301), we found that most users (76.4%, 230/301) were still engaging with the app at least once every 30 days.

Rates of nonuse attrition in this study were low (<40%) compared with previous research and commercial apps. A similar RCT found that 80% of participants ceased using a Web-based PA intervention by week 80 [28], although this study duration is substantially longer than this study. In the Web-based 10,000 steps Australia study [29], 74.84% (8720/11,651) of app-only users had stopped using the intervention by day 43, although this higher rate is because the study occurred in a real world rather than controlled setting. Another real-world study of the 10,000 steps Australia website found that nonuse attrition occurred for 78% of participants after just 2 weeks [22]. Note that each of these studies [22,28,29] used 14 days of nonuse as their threshold, which is likely to have inflated their attrition rate compared with the 30-day threshold we used. When we applied the 14-day threshold in the sensitivity analysis, the attrition rate increased to 50% for the gamified group and 59% for the basic group across the 100-day period, which still compares favorably to the 10,000 steps studies [22,29]. However, our rate of attrition compared favorably with commercial apps, used in real-world settings, which experience an average rate of 62% [35] and which use the 30 consecutive days of nonuse threshold. That nonuse attrition was low in this study, and in that by Kolt et al [28], may be because of the fact that participants were using the app within the context of an RCT [36,37], compared with someone using an app under more ecologically valid circumstances. From our results, we cannot comment on engagement or attrition beyond the 100-day intervention period, as previous research reports increasingly rapid decline of usage as time passes [8,9,38-40].

Consistent with previous research [11], participant characteristics, namely being older and not being obese, were associated with increased engagement. Active Team was designed to be simple and easy to use, and it may have appealed to older users who are potentially less savvy in their use of technology or who may be more health conscious, as they tend to face more health problems [41-43]. Other intervention studies [29,44,45] have found higher engagement among older users, and a systematic review [11] reported a trend toward age and engagement being positively correlated. At first, the fact that users with higher BMI engaged with the app less is discouraging, and other studies have reported a similar negative association [11], as this suggests people who would benefit from the program the most were the least engaged, which is a common challenge for health promoters, whereby groups with the highest needs tend to be the hardest to recruit and engage [46,47]. However, in this study, a high percentage of overweight and obese users (78.4%, 236/301) compared with approximately 63% of the population) [48] were recruited; however, the engagement data show these users engaged less across the intervention period, perhaps as risks associated with PA, such as joint and respiratory problems [49-51], become more pronounced for people with higher BMIs, which does make engaging in PA more difficult. In contrast, an intervention using Fitbits found that overweight participants were more likely to engage and increase their PA [52], and more work should be undertaken to understand how best to engage different groups. Regardless, it should be noted that PA holds many health benefits beyond the control of body weight, including improved mood [53,54] and cardiorespiratory fitness [55], which means the higher engagement reported by participants who reported a BMI in the healthy or overweight range is still beneficial.

Users of the gamified app had higher rates of engagement (as measured by each of our metrics), and they were more likely to be super users (74%, 56/76 of super users had the gamified app). This suggests the social and gamified features enhanced engagement [56,57]. In this study, this is consistent with our intent at the outset, as the social and gamified features and prompts were specifically designed and included to increase engagement, these users also received push notifications when someone interacted with them in the app, which were intended to draw them back to the app. As all of our gamified features were social in nature, this network effect likely influenced rates of app use; use by 1 team member (or lack thereof) may have had a flow on effect, through the team, which either promoted or inhibited use. Although our current analysis does not allow us to comment on which social and gamified features resulted in increased engagement, our results suggest that future apps could benefit from including gamified features to promote comparison and competition among users. Moreover, some reviews [58,59] suggest the inclusion of more sophisticated gamified features (eg, personalized avatars, storylines) could enhance usage further. This should be considered with the caveat that there is ongoing debate around whether external motivation provided by gamification undermines intrinsic motivation for participating in PA [60-64]. This is particularly true for the gamified Active Team app, where continued usage may be dependent on usage by others, which could further undermine the intrinsic motivation. Although gamification does need to be implemented carefully to support establishment of and commitment to a PA regimen, the provision of an extrinsic motivator (to begin) is preferable to lacking any motivation.

Despite similar baseline PA levels, super users saw statistically significant increases in their objectively measured PA over time, when compared with regular users. These results held true after controlling for demographic characteristics likely to influence engagement (ie, age, BMI, sex, and education level [34]). Although promising, this does mean that regular users, who made up the majority of users (74.8%, (225/301), had PA levels that, on average, decreased slightly, 8 (SD 47) min, over the 3-month period. Taken together, these results suggest that

app-based interventions can work, perhaps only for a small group of people or only when engagement can be maximized [8,65]. Although users of the gamified app had more features to engage with, these results are near identical when engagement was assessed using step calendar use, a feature that both groups had access to (see Multimedia Appendix 1). This suggests that self-monitoring of behavior may have been an engaging and potent enough mechanism for change for some users [66].

Together, these findings highlight the complexities of the theorized dose-response relationship between intervention exposure (engagement) and efficacy [8-10,65] and support the theory of an optimum dosage or exposure *threshold* [67] required for intervention efficacy, which is potentially different for each user. Given super users saw statistically significant increases to their PA over time, whereas those who used the app less did not, also suggests ongoing efforts to understand and boost engagement [12,13,68] are warranted to enhance appeal so that as many users as possible receive the optimum exposure for intervention benefits.

Strengths, Limitations, and Future Directions

The findings presented here should be interpreted in the context of the study's limitations. First, only user engagement during the 100-day intervention duration is included; we have relied on objective engagement metrics from the study's server, and we have used the term gamification to refer to a particular set of game design elements (ie, newsfeed, challenges, virtual gifts, and leaderboard), all of which are social in nature. Furthermore, this study has not considered psychological aspects of engagement, such as participant perceptions or the interplay between personality and engagement [12,40]. In this study, super users were *a priori* defined as those whose app use fell within the top quartile of all users; however, we acknowledge that other definitions of this novel term would have been equally justified. Moreover, no single engagement metric can wholly capture and explain user engagement, as such, there will be limitations associated with any engagement metric. For example, the concept of nonuse attrition may be blunt, as it does fail to adequately capture usage that diminishes over time in those who continue to engage with the program. Our choice of engagement metric was guided by the metrics reported by industry-led apps, and it is important to note that such apps often view uptake and use as benchmark for success rather than behavior change. However, in research and public health settings, engagement is considered important, as it mediates user exposure to the intervention and therefore has important implications for efficacy. Causal claims cannot be made, as this study is a secondary and subgroup analysis of RCT data that did not experimentally test the impact of varying levels of engagement on PA outcomes. Reverse causation is equally possible; the relationships between app usage and PA may be because users who were able to increase their PA over time were more satisfied with their experience of the intervention; therefore, they used the app more. The sample is also subject to a self-selection bias and, typical of health behavior research, well-educated women are overrepresented [46].

Strengths of this study include the large sample size (n= 301) and the inclusion of objective measurements of app usage and PA. Although we acknowledge participants signing up to an RCT may be highly motivated, our intervention was delivered entirely via an app with no in-person contact between study staff and participants. However, the inclusion of eligibility criteria, baseline, and 3-month follow-up assessments signifies this approach is distinct from real-world settings, and we do not know whether our high rates of engagement are potentially influenced by a Hawthorne effect (where awareness of being under observation results in modified behavior [69,70]). Translation of researcher-led apps into real-world trials is likely to become more commonplace [7,21,22,71], and the ability to make comparisons to available industry-led apps will be increasingly important. This study has used engagement metrics commonly reported by industry-led apps, and the method and findings may be useful to future studies seeking to make direct comparisons between research- and industry-led apps.

A real-world trial of the app used in this study is now underway, and once complete, it can confirm whether the findings reported here hold true in a setting comparable to how we would expect people to use commercially available apps. Our findings support the inclusion of social and gamified features to increase engagement. Interventions are increasingly incorporating gamification and social features [45,72-75], and it will be interesting to see whether our engagement results are replicated in these studies. Moreover, although consensus on how engagement is operationalized has not yet eventuated, the field of engagement science can continue to progress if future studies publish analyses of app engagement that are similarly, or more, detailed (ie, to consider each gamified feature separately) than this study. These analyses should seek to evaluate the role of engagement levels on intervention efficacy in both instances where interventions are able to demonstrate overall efficacy and those that do not. In this way, engaging features can be identified, refined, and incorporated into future interventions to enhance their potential for positive behavior change effects.

Conclusions

Taken together, our results indicate app-based interventions can sustain user engagement across a 100-day intervention period, and the inclusion of social and gamified features can enhance engagement. Users who were older or with a BMI classified as healthy or overweight (rather than obese) engaged more, and those who engaged the most (ie, super users, the top quartile of users) reported statistically significant increases to their objectively measured PA over time, supporting the theory that intervention exposure is linked to efficacy. Further research is now needed to determine whether these findings are replicated in other app- and Web-based behavior change studies and in ecologically valid real-world settings.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Step calendar use.
[PDF File (Adobe PDF File), 406 KB - jmir_v21i11e14645_app1.pdf]

References

1. Kohl LF, Crutzen R, de Vries NK. Online prevention aimed at lifestyle behaviors: a systematic review of reviews. J Med Internet Res 2013 Jul 16;15(7):e146 [FREE Full text] [doi: 10.2196/jmir.2665] [Medline: 23859884]
2. Lee M, Lee H, Kim Y, Kim J, Cho M, Jang J, et al. Mobile app-based health promotion programs: a systematic review of the literature. Int J Environ Res Public Health 2018 Dec 13;15(12):pii: E2838 [FREE Full text] [doi: 10.3390/ijerph15122838] [Medline: 30551555]
3. Müller AM, Maher CA, Vandelanotte C, Hingle M, Middelweerd A, Lopez ML, et al. Physical activity, sedentary behavior, and diet-related eHealth and mHealth research: bibliometric analysis. J Med Internet Res 2018 Apr 18;20(4):e122 [FREE Full text] [doi: 10.2196/jmir.8954] [Medline: 29669703]
4. Maher C, Ferguson M, Vandelanotte C, Plotnikoff R, de Bourdeaudhuij I, Thomas S, et al. A web-based, social networking physical activity intervention for insufficiently active adults delivered via Facebook app: randomized controlled trial. J Med Internet Res 2015 Jul 13;17(7):e174 [FREE Full text] [doi: 10.2196/jmir.4086] [Medline: 26169067]
5. Wing RR, Pinto AM, Crane MM, Kumar R, Weinberg BM, Gorin AA. A statewide intervention reduces BMI in adults: Shape Up Rhode Island results. Obesity (Silver Spring) 2009 May;17(5):991-995 [FREE Full text] [doi: 10.1038/oby.2008.655] [Medline: 19180068]
6. Gal R, May AM, van Overmeeren EJ, Simons M, Monninkhof EM. The effect of physical activity interventions comprising wearables and smartphone applications on physical activity: a systematic review and meta-analysis. Sports Med Open 2018 Sep 3;4(1):42 [FREE Full text] [doi: 10.1186/s40798-018-0157-9] [Medline: 30178072]
7. Romeo A, Edney S, Plotnikoff R, Curtis R, Ryan J, Sanders I, et al. Can smartphone apps increase physical activity? Systematic review and meta-analysis. J Med Internet Res 2019 Mar 19;21(3):e12053 [FREE Full text] [doi: 10.2196/12053] [Medline: 30888321]
8. Maher CA, Lewis LK, Ferrar K, Marshall S, de Bourdeaudhuij I, Vandelanotte C. Are health behavior change interventions that use online social networks effective? A systematic review. J Med Internet Res 2014 Feb 14;16(2):e40 [FREE Full text] [doi: 10.2196/jmir.2952] [Medline: 24550083]
9. Laranjo L, Arguel A, Neves AL, Gallagher AM, Kaplan R, Mortimer N, et al. The influence of social networking sites on health behavior change: a systematic review and meta-analysis. J Am Med Inform Assoc 2015 Jan;22(1):243-256 [FREE Full text] [doi: 10.1136/amiajnl-2014-002841] [Medline: 25005606]
10. Cavallo DN, Tate DF, Ward DS, DeVellis RF, Thayer LM, Ammerman AS. Social support for physical activity-role of Facebook with and without structured intervention. Transl Behav Med 2014 Dec;4(4):346-354 [FREE Full text] [doi: 10.1007/s13142-014-0269-9] [Medline: 25584083]
11. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
12. Short CE, DeSmet A, Woods C, Williams SL, Maher C, Middelweerd A, et al. Measuring engagement in eHealth and mHealth behavior change interventions: viewpoint of methodologies. J Med Internet Res 2018 Nov 16;20(11):e292 [FREE Full text] [doi: 10.2196/jmir.9397] [Medline: 30446482]
13. Pham Q, Graham G, Carrion C, Morita PP, Seto E, Stinson JN, et al. A library of analytic indicators to evaluate effective engagement with consumer mHealth apps for chronic conditions: scoping review. JMIR Mhealth Uhealth 2019 Jan 18;7(1):e11941 [FREE Full text] [doi: 10.2196/11941] [Medline: 30664463]
14. Dennison L, Morrison L, Conway G, Yardley L. Opportunities and challenges for smartphone applications in supporting health behavior change: qualitative study. J Med Internet Res 2013 Apr 18;15(4):e86 [FREE Full text] [doi: 10.2196/jmir.2583] [Medline: 23598614]
15. Bonnie E. CleverTap. 2017. Churn Rate: How to Define and Calculate Customer Churn. URL: https://clevertap.com/blog/churn-rate/ [accessed 2019-07-25] [WebCite Cache ID 78AG7rOi8]
16. Global App Testing. 2019. 6 Secrets of App Stickiness. URL: https://www.globalapptesting.com/blog/the-6-secrets-of-app-stickiness [accessed 2019-07-25] [WebCite Cache ID 78DD3WxDJ]
17. NewGenApps. 2017. 10 Key Metrics to Measure User Engagement in Mobile Apps. URL: https://www.newgenapps.com/blog/mobile-app-analytics-10-key-metrics-to-measure-user-engagement [accessed 2019-06-25] [WebCite Cache ID 78FhtZC4x]
18. Valdellon L. CleverTap. 2018. How to Measure Your App's Health with User Engagement Metrics. URL: https://clevertap.com/blog/measure-app-health-with-user-engagement-metrics/ [accessed 2019-06-25] [WebCite Cache ID 78Fi6zlSv]
19. Rojek R. Influencive. 2017. Recognize Your Super Users and Accelerate Your App Growth. URL: https://www.influencive.com/recognize-super-users-accelerate-app-growth/ [accessed 2019-06-25] [WebCite Cache ID 78DDLGl33]

20. Vandelanotte C, Müller AM, Short CE, Hingle M, Nathan N, Williams SL, et al. Past, present, and future of eHealth and mHealth research to improve physical activity and dietary behaviors. J Nutr Educ Behav 2016 Mar;48(3):219-228.e1. [doi: 10.1016/j.jneb.2015.12.006] [Medline: 26965100]
21. Vandelanotte C, Maher CA. Why we need more than just randomized controlled trials to establish the effectiveness of online social networks for health behavior change. Am J Health Promot 2015;30(2):74-76. [doi: 10.4278/ajhp.141204-CIT-605] [Medline: 26517586]
22. Vandelanotte C, Kolt GS, Caperchione CM, Savage TN, Rosenkranz RR, Maeder AJ, et al. Effectiveness of a web 2.0 intervention to increase physical activity in real-world settings: randomized ecological trial. J Med Internet Res 2017 Nov 13;19(11):e390 [FREE Full text] [doi: 10.2196/jmir.8484] [Medline: 29133282]
23. Edney S, Plotnikoff R, Vandelanotte C, Olds T, de Bourdeaudhuij I, Ryan J, et al. "Active Team" a social and gamified app-based physical activity intervention: randomised controlled trial study protocol. BMC Public Health 2017 Nov 2;17(1):859 [FREE Full text] [doi: 10.1186/s12889-017-4882-7] [Medline: 29096614]
24. Deterding S, Dixon D, Khaled R, Nacke L. From Game Design Elements to Gamefulness: Defining 'Gamification'. In: MindTrek'11. 2011 Presented at: Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments; September 28-30, 2011; Tampere, Finland p. 9-15. [doi: 10.1145/2181037.2181040]
25. Esliger DW, Rowlands AV, Hurst TL, Catt M, Murray P, Eston RG. Validation of the GENEA accelerometer. Med Sci Sports Exerc 2011 Jun;43(6):1085-1093. [doi: 10.1249/MSS.0b013e31820513be] [Medline: 21088628]
26. Australian Institute of Health and Welfare. The Active Australia Survey: A Guide and Manual for Implementation, Analysis and Reporting. Canberra: AIHW; 2003.
27. Iqbal M. Business of Apps. 2019. Pokémon GO Revenue and Usage Statistics (2019). URL: https://www.businessofapps.com/data/pokemon-go-statistics/ [accessed 2019-06-30]
28. Kolt GS, Rosenkranz RR, Vandelanotte C, Caperchione CM, Maeder AJ, Tague R, et al. Using Web 2.0 applications to promote health-related physical activity: findings from the WALK 2.0 randomised controlled trial. Br J Sports Med 2017 Oct;51(19):1433-1440 [FREE Full text] [doi: 10.1136/bjsports-2016-096890] [Medline: 28049624]
29. Guertler D, Vandelanotte C, Kirwan M, Duncan MJ. Engagement and nonusage attrition with a free physical activity promotion program: the case of 10,000 steps Australia. J Med Internet Res 2015 Jul 15;17(7):e176 [FREE Full text] [doi: 10.2196/jmir.4339] [Medline: 26180040]
30. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. J Am Stat Assoc 1958 Jun;53(282):457-481. [doi: 10.1080/01621459.1958.10501452]
31. Serrano KJ, Coa KI, Yu M, Wolff-Hughes DL, Atienza AA. Characterizing user engagement with health app data: a data mining approach. Transl Behav Med 2017 Jun;7(2):277-285 [FREE Full text] [doi: 10.1007/s13142-017-0508-y] [Medline: 28616846]
32. Statista. 2019. Percentage of Mobile Apps That Have Been Used Only Once During the First Six Months of Ownership From 2010 to 2018. URL: https://www.statista.com/statistics/271628/percentage-of-apps-used-once-in-the-us/ [accessed 2019-06-30] [WebCite Cache ID 78DDkCwca]
33. Thrul J, Klein AB, Ramo DE. Smoking cessation intervention on Facebook: which content generates the best engagement? J Med Internet Res 2015 Nov 11;17(11):e244 [FREE Full text] [doi: 10.2196/jmir.4575] [Medline: 26561529]
34. Ryan J, Edney S, Maher C. Engagement, compliance and retention with a gamified online social networking physical activity intervention. Transl Behav Med 2017 Dec;7(4):702-708 [FREE Full text] [doi: 10.1007/s13142-017-0499-8] [Medline: 28523603]
35. Rodde T. Localytics. 2018. 21% of Users Abandon an App After One Use. URL: http://info.localytics.com/blog/21-percent-of-users-abandon-apps-after-one-use [accessed 2019-06-30] [WebCite Cache ID 78DDPGwEb]
36. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
37. Kelders SM, Kok RN, Ossebaard HC, van Gemert-Pijnen JE. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012 Nov 14;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]
38. Shin G, Jarrahi MH, Fei Y, Karami A, Gafinowitz N, Byun A, et al. Wearable activity trackers, accuracy, adoption, acceptance and health impact: a systematic literature review. J Biomed Inform 2019 May;93:103153. [doi: 10.1016/j.jbi.2019.103153] [Medline: 30910623]
39. Clawson J, Pater J, Miller A, Mynatt E, Mamykina L. No Longer Wearing: Investigating the Abandonment of Personal Health-tracking Technologies on Craigslist. In: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing. 2015 Presented at: UbiComp'15; September 7-11, 2015; Osaka, Japan p. 07-11. [doi: 10.1145/2750858.2807554]
40. Maher C, Ryan J, Ambrosi C, Edney S. Users' experiences of wearable activity trackers: a cross-sectional study. BMC Public Health 2017 Nov 15;17(1):880 [FREE Full text] [doi: 10.1186/s12889-017-4888-1] [Medline: 29141607]
41. Venkatesh V, Thong JY, Xu X. Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. MIS Quarterly 2012;36(1):157-178. [doi: 10.2307/41410412]

42. Carpenter BD, Buday S. Computer use among older adults in a naturally occurring retirement community. Comp Hum Behav 2007 Nov;23(6):3012-3024 [FREE Full text] [doi: 10.1016/j.chb.2006.08.015]
43. Nigg CR, Burbank PM, Padula C, Dufresne R, Rossi JS, Velicer WF, et al. Stages of change across ten health risk behaviors for older adults. Gerontologist 1999 Aug;39(4):473-482. [doi: 10.1093/geront/39.4.473] [Medline: 10495586]
44. Davies C, Corry K, van Itallie A, Vandelanotte C, Caperchione C, Mummery WK. Prospective associations between intervention components and website engagement in a publicly available physical activity website: the case of 10,000 Steps Australia. J Med Internet Res 2012 Jan 11;14(1):e4 [FREE Full text] [doi: 10.2196/jmir.1792] [Medline: 22260810]
45. Kappen D, Mirza-Babaei P, Nacke L. Gamification of Older Adults' Physical Activity: An Eight-Week Study. In: Proceedings of the 51st Hawaii International Conference on System Sciences. 2018 Presented at: HICSS'18; January 3-6, 2018; Hawaii URL: http://hdl.handle.net/10125/50036 [doi: 10.24251/hicss.2018.149]
46. Plotnikoff RC, Costigan SA, Williams RL, Hutchesson MJ, Kennedy SG, Robards SL, et al. Effectiveness of interventions targeting physical activity, nutrition and healthy weight for university and college students: a systematic review and meta-analysis. Int J Behav Nutr Phys Act 2015 Apr 1;12:45 [FREE Full text] [doi: 10.1186/s12966-015-0203-7] [Medline: 25890337]
47. Bonevski B, Randell M, Paul C, Chapman K, Twyman L, Bryant J, et al. Reaching the hard-to-reach: a systematic review of strategies for improving health and medical research with socially disadvantaged groups. BMC Med Res Methodol 2014 Mar 25;14:42 [FREE Full text] [doi: 10.1186/1471-2288-14-42] [Medline: 24669751]
48. Australian Institute of Health and Welfare. Australia's Health 2018. Australia: AIHW; 2018.
49. Berg FM. Health risks associated with weight loss and obesity treatment programs. J Soc Issues 2002 Dec 17;55(2):277-297 [FREE Full text] [doi: 10.1111/0022-4537.00116]
50. Bazelmans C, Coppieters Y, Godin I, Parent F, Berghmans L, Dramaix M, et al. Is obesity associated with injuries among young people? Eur J Epidemiol 2004;19(11):1037-1042. [doi: 10.1007/s10654-004-0158-5] [Medline: 15648597]
51. Sallinen J, Leinonen R, Hirvensalo M, Lyyra T, Heikkinen E, Rantanen T. Perceived constraints on physical exercise among obese and non-obese older people. Prev Med 2009 Dec;49(6):506-510. [doi: 10.1016/j.ypmed.2009.10.001] [Medline: 19833148]
52. Vandelanotte C, Duncan MJ, Maher CA, Schoeppe S, Rebar AL, Power DA, et al. The effectiveness of a web-based computer-tailored physical activity intervention using fitbit activity trackers: randomized trial. J Med Internet Res 2018 Dec 18;20(12):e11321 [FREE Full text] [doi: 10.2196/11321] [Medline: 30563808]
53. Annesi JJ, Porter KJ, Hill GM, Goldfine BD. Effects of instructional physical activity courses on overall physical activity and mood in university students. Res Q Exerc Sport 2017 Sep;88(3):358-364. [doi: 10.1080/02701367.2017.1336280] [Medline: 28636503]
54. Penedo FJ, Dahn JR. Exercise and well-being: a review of mental and physical health benefits associated with physical activity. Curr Opin Psychiatry 2005 Mar;18(2):189-193. [doi: 10.1097/00001504-200503000-00013] [Medline: 16639173]
55. Warburton DE, Nicol CW, Bredin SS. Health benefits of physical activity: the evidence. Can Med Assoc J 2006 Mar 14;174(6):801-809 [FREE Full text] [doi: 10.1503/cmaj.051351] [Medline: 16534088]
56. Looyestyn J, Kernot J, Boshoff K, Ryan J, Edney S, Maher C. Does gamification increase engagement with online programs? A systematic review. PLoS One 2017;12(3):e0173403 [FREE Full text] [doi: 10.1371/journal.pone.0173403] [Medline: 28362821]
57. Patel MS, Benjamin EJ, Volpp KG, Fox CS, Small DS, Massaro JM, et al. Effect of a game-based intervention designed to enhance social incentives to increase physical activity among families: the Be Fit randomized clinical trial. JAMA Intern Med 2017 Nov 1;177(11):1586-1593 [FREE Full text] [doi: 10.1001/jamainternmed.2017.3458] [Medline: 28973115]
58. Sardi L, Idri A, Fernández-Alemán JL. A systematic review of gamification in e-Health. J Biomed Inform 2017 Jul;71:31-48 [FREE Full text] [doi: 10.1016/j.jbi.2017.05.011] [Medline: 28536062]
59. Johnson D, Deterding S, Kuhn K, Staneva A, Stoyanov S, Hides L. Gamification for health and wellbeing: a systematic review of the literature. Internet Interv 2016 Nov;6:89-106 [FREE Full text] [doi: 10.1016/j.invent.2016.10.002] [Medline: 30135818]
60. Deci EL, Koestner R, Ryan RM. A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. Psychol Bull 1999 Nov;125(6):627-68; discussion 692. [doi: 10.1037/0033-2909.125.6.627] [Medline: 10589297]
61. Suh A, Wagner C, Liu L. Enhancing user engagement through gamification. J Comput Inform Syst 2016 Oct 3;58(3):204-213. [doi: 10.1080/08874417.2016.1229143]
62. Attig C, Franke T. I track, therefore I walk – exploring the motivational costs of wearing activity trackers in actual users. Int J Hum-Comput St 2019 Jul;127:211-224. [doi: 10.1016/j.ijhcs.2018.04.007]
63. Kyewski E, Krämer NC. To gamify or not to gamify? An experimental field study of the influence of badges on motivation, activity, and performance in an online learning course. Comput Educ 2018 Mar;118:25-37. [doi: 10.1016/j.compedu.2017.11.006]
64. Murray JM, French DP, Patterson CC, Kee F, Gough A, Tang J, et al. Predicting outcomes from engagement with specific components of an internet-based physical activity intervention with financial incentives: process analysis of a cluster

randomized controlled trial. J Med Internet Res 2019 Apr 19;21(4):e11394 [FREE Full text] [doi: 10.2196/11394] [Medline: 31002304]
65. Vandelanotte C, Spathonis KM, Eakin EG, Owen N. Website-delivered physical activity interventions a review of the literature. Am J Prev Med 2007 Jul;33(1):54-64. [doi: 10.1016/j.amepre.2007.02.041] [Medline: 17572313]
66. Samdal GB, Eide GE, Barth T, Williams G, Meland E. Effective behaviour change techniques for physical activity and healthy eating in overweight and obese adults; systematic review and meta-regression analyses. Int J Behav Nutr Phys Act 2017 Mar 28;14(1):42 [FREE Full text] [doi: 10.1186/s12966-017-0494-y] [Medline: 28351367]
67. Nelson LA, Coston TD, Cherrington AL, Osborn CY. Patterns of user engagement with mobile- and web-delivered self-care interventions for adults with T2DM: a review of the literature. Curr Diab Rep 2016 Jul;16(7):66 [FREE Full text] [doi: 10.1007/s11892-016-0755-1] [Medline: 27255269]
68. Danaher BG, Brendryen H, Seeley JR, Tyler MS, Woolley T. From black box to toolbox: outlining device functionality, engagement activities, and the pervasive information architecture of mHealth interventions. Internet Interv 2015 Mar 1;2(1):91-101 [FREE Full text] [doi: 10.1016/j.invent.2015.01.002] [Medline: 25750862]
69. Wickström G, Bendix T. The 'Hawthorne effect'--what did the original Hawthorne studies actually show? Scand J Work Environ Health 2000 Aug;26(4):363-367 [FREE Full text] [doi: 10.5271/sjweh.555] [Medline: 10994804]
70. Sedgwick P, Greenwood N. Understanding the Hawthorne effect. Br Med J 2015 Sep 4;351:h4672. [doi: 10.1136/bmj.h4672] [Medline: 26341898]
71. Vandelanotte C, Duncan MJ, Kolt GS, Caperchione CM, Savage TN, van Itallie A, et al. More real-world trials are needed to establish if web-based physical activity interventions are effective. Br J Sports Med 2018 Jul 3:- (forthcoming). [doi: 10.1136/bjsports-2018-099437] [Medline: 29970409]
72. Boekhout JM, Peels DA, Berendsen BA, Bolman CA, Lechner L. An eHealth intervention to promote physical activity and social network of single, chronically impaired older adults: adaptation of an existing intervention using intervention mapping. JMIR Res Protoc 2017 Nov 23;6(11):e230 [FREE Full text] [doi: 10.2196/resprot.8093] [Medline: 29170146]
73. Lee D, Frey G, Cheng A, Shih PC. Puzzle Walk: A Gamified Mobile App to Increase Physical Activity in Adults with Autism Spectrum Disorder. In: Proceedings of the 10th International Conference on Virtual Worlds and Games for Serious Applications. 2018 Presented at: VS-Games'18; September 5-7, 2018; Wurzburg, Germany. [doi: 10.1109/vs-games.2018.8493439]
74. Corepal R, Best P, O'Neill R, Tully MA, Edwards M, Jago R, et al. Exploring the use of a gamified intervention for encouraging physical activity in adolescents: a qualitative longitudinal study in Northern Ireland. BMJ Open 2018 Apr 20;8(4):e019663 [FREE Full text] [doi: 10.1136/bmjopen-2017-019663] [Medline: 29678971]
75. van Woudenberg TJ, Bevelander KE, Burk WJ, Smit CR, Buijs L, Buijzen M. A randomized controlled trial testing a social network intervention to promote physical activity among adolescents. BMC Public Health 2018 Apr 23;18(1):542 [FREE Full text] [doi: 10.1186/s12889-018-5451-4] [Medline: 29685112]

Abbreviations

AAS: Active Australia Survey
BMI: body mass index
DAU: daily active user
MAU: monthly active user
MVPA: moderate-to-vigorous physical activity
PA: physical activity
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 09.05.19; peer-reviewed by S Baptista, M Patel, T Cahill; comments to author 21.06.19; revised version received 30.07.19; accepted 30.08.19; published 27.11.19.

Please cite as:
Edney S, Ryan JC, Olds T, Monroe C, Fraysse F, Vandelanotte C, Plotnikoff R, Curtis R, Maher C
User Engagement and Attrition in an App-Based Physical Activity Intervention: Secondary Analysis of a Randomized Controlled Trial
J Med Internet Res 2019;21(11):e14645
URL: http://www.jmir.org/2019/11/e14645/
doi:10.2196/14645
PMID:31774402

Original Paper

Patients' Use of the Internet to Find Reliable Medical Information About Minor Ailments: Vignette-Based Experimental Study

Joyce Kwakernaak[1], MD; Just A H Eekhof[1], MD, PhD; Margot W M De Waal[1], MSc, PhD; Elisabeth A M Barenbrug[2], MSc; Niels H Chavannes[1], MD

[1]Department Public Health & Primary Care, Leiden University Medical Centre, Leiden, Netherlands

[2]Consumentenbond, Den Haag, Netherlands

Corresponding Author:
Just A H Eekhof, MD, PhD
Department Public Health & Primary Care
Leiden University Medical Centre
PO Box 9600
Leiden, 2300RC
Netherlands
Phone: 31 715268414
Email: j.a.h.eekhof@lumc.nl

Abstract

Background: Little is known about the exact process of how patients search for medical information on the internet and what they retrieve. There is especially a paucity of literature on browsing for information on minor ailments, a term used for harmless diseases that are very common in the general population and thus have a significant impact on health care.

Objective: This vignette-based experimental study aimed to explore what kind of Web-based search strategies are applied and how search strategies, demographic characteristics, and the quality of the visited websites relate to finding the right diagnosis. Additional goals were to describe how searching on the Web influences one's perception of the severity of the potential diagnosis and whether or not the participants would discuss the information they found on the internet with their doctors.

Methods: Out of 1372 survey participants, 355 were randomly sampled, and 155 of them were recruited and assigned to one of four clinical scenarios. Each search term they used was classified as one of three search strategies: (1) hypothesis testing, (2) narrowing within the general hypothesis area, and (3) symptom exploration. The quality of the websites used was determined by using the DISCERN instrument. To compare the diagnostic accuracy of the participants before and after the internet search, a McNemar test was used. Chi-square tests were used to describe which factors are related to the chosen search strategy. A multivariate binary logistic regression model was constructed to predict which factors are related to finding a sound diagnosis after searching the internet for health information.

Results: Most participants (65.8%, 102/155) used the symptom exploration strategy. However, this depends on the assigned scenario (P<.001) and the self-estimated severity score of the symptoms before the internet search (P=.001). A significant relation was found between choosing an accurate diagnosis and age (odds ratio [OR] 0.94, 95% CI 0.90 to 0.98) and the clinical scenario, as well as the use of high-quality websites (OR 7.49, 95% CI 1.85 to 30.26). Browsing the internet did not lead to a statistically significant change in participants' beliefs about the severity of the condition (McNemar test, P=.85). Most participants (65%) shared their retrieved information with their physician and most of them (75%) received a positive response.

Conclusions: Our findings suggest that most patients use a symptom-based approach; however, if patients expect the potential diagnosis to be severe, they tend to use a hypothesis verification strategy more often and are therefore prone to certain forms of bias. In addition, self-diagnosing accuracy is related to younger age, the symptom scenario, and the use of high-quality websites. We should find ways to guide patients toward search strategies and websites that may more likely lead to accurate decision making.

(J Med Internet Res 2019;21(11):e12278) doi:10.2196/12278

KEYWORDS

internet; information seeking behaviour; consumer health information; diagnosis; humans; adult

Introduction

Background

Over the last few decades, the internet has become an important and easily accessible source of medical information for patients [1,2]. A reason that people frequently mention for visiting the Web is to find reassurance or to find an explanation or diagnosis for their physical complaints [2]. Furthermore, people tend to use the internet for determining whether or not they should consult a physician [2]. Younger patients, females, and highly educated individuals are known to search for health information more often than others [3,4]. Internet use has the potential to improve patient empowerment, as well-informed patients tend to play a more active role in their own health care [5]. In addition, it may influence the patients' timing of help seeking behavior [6]. These findings also correspond with the results of a study performed on the widely used website, Thuisarts.nl. This is a reliable source of medical information about minor ailments and advice on self-care, developed by the Dutch College of General Practitioners. Research has shown that in the 2 years after the introduction of this website, the number of short general practitioner (GP) consultations decreased by 12% [7]. However, there are downsides to the use of this easily accessible source of information. It should be considered that most patients are not medically trained. This makes it difficult for them to understand medical jargons and to find accurate information, especially for those with lower education levels and a low socioeconomic status [8,9]. Furthermore, there is an abundance of low-quality websites available, as only a minority meets the standardized quality and accuracy requirements [10,11]. However, no less than two-thirds of the patients have a high level of confidence in the information retrieved from the internet [12]. Using this incorrect information can lead to incorrect decisions [13,14] or to the retrieval of incorrect notions about a medical condition or treatment [15,16]. In addition, lay individuals are often inaccurate when they try to self-diagnose without consulting a physician [17].

Previous Research

What remains relatively unexplored in the current literature is exactly how patients look for information on the internet. A certain approach was chosen by Pang et al [18], who divided search behaviors into 4 different categories. The distinction was not made to determine the best strategy but to indicate that the following 4 different search strategies have different needs in their search for information: *Quick Fact Seeking* refers to patients terminating their search once they retrieved superficial information for a specific health issue. Therefore, websites should provide key points and a brief summary that is relevant to the topic; *All-Around Skimming* indicates patients who go through a wide range of information in a fast manner. To support this behavior, excerpts and previews will be helpful; *Focused Reading* denotes concentrated reading on a particular topic. Reader-friendly features are recommended to support this behavior; and *Knowledge Digging* indicates the intense reading associated with the in-depth research on a number of diverse health topics. Therefore, a broader range of information should be provided. They created a design for consumer health websites that meets these different needs and concluded that this approach will lead to better knowledge acquisition. Other literature on this topic has been derived from experimental studies, where participants were assigned to search the internet for information on hypothetical scenarios. Keselman et al [19] observed that lay individuals using the *verification of the primary hypothesis* strategy tend to seek out data that correspond to their incorrect initial diagnosis to confirm their own hypothesis and health beliefs (confirmation bias). Participants using the *narrowing search within the general hypothesis* strategy remained indecisive about the diagnosis, whereas the *bottom-up symptom exploration* strategy seemed the most successful. Luger et al [20] conducted a study that revealed that using previous illness experiences and having less existing medical knowledge were associated with choosing an inaccurate diagnosis. Perez et al [21,22] explain that their results correspond with the dual-processing theory, differentiating between system 1 processing (unconscious, initiative, automatic, rapid, low effort) and system 2 processing (conscious, systematic, deliberative, slow, high effort). System 2 processing is associated with higher-quality decision making and is more often applied by individuals with a higher education level and younger age. Another previous study found that patients often select websites of organizations that they consider to be of good reputation or organizations that are domestic, because it makes them feel more confident that they are getting reliable information. On the contrary, participants often avoid websites that have visible advertising or are obviously profit-oriented [23].

Study Aim

To address the challenging process of obtaining and applying medical information derived from the internet by patients, it is of great importance to understand how these individuals search for information. There is, however, limited information available on this topic. So far, there has been no research into how people search for medical information on the internet about minor ailments, whereas these are common health issues affecting a large part of everyday health care in general practice. The aim of this study was to explore what kind of Web-based search strategies are being used and if they lead to finding a sound diagnosis. Additional objectives are to determine whether the quality of the used websites and certain participant characteristics are related to finding the right diagnosis. Furthermore, this paper describes how searching on the Web influences the perception of severity of the potential diagnosis and if the participants in general would discuss the retrieved information on the internet with their doctor.

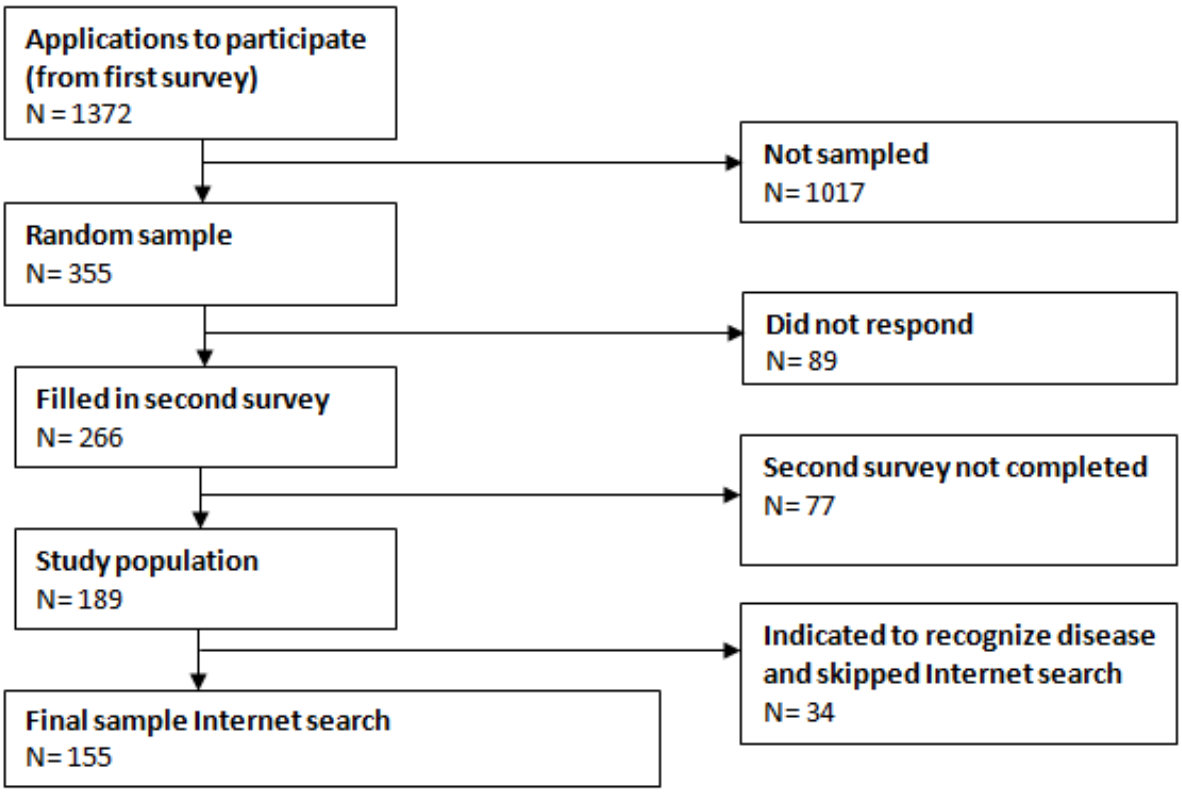
Methods

Recruitment of Participants

This research was conducted on a vignette-based experimental study design that focused on the internet search patterns of Dutch adults. Participants were recruited through a survey on the use of the internet for obtaining information about medical issues, which can be found in Multimedia Appendix 1. This survey was compiled in collaboration with the Dutch Consumers' Association, which is an independent nonprofit association that conducts research and makes publications about various products, services, and injustice in society. The survey

was sent to the Association's panel members who could fill out the Web-based survey from January 18 to January 27, 2017. This resulted in a total number of 5774 respondents. These participants were asked if they were willing to apply for the experimental test. However, they were only able to do so if they met the inclusion and exclusion criteria. The inclusion criteria were that the participants have access to the internet, that they use it for obtaining information about medical issues, and that they were aged 18 years or older. In addition, the participants were not able to apply if they or their housemates were medically educated. A random sample was taken from the resulting 1372 applicants, which resulted in 355 selected individuals who received an invitation email. In total, 189 of the 355 selected participants were included in this study, of which 155 completed the internet search. The 34 participants who did not complete the internet search indicated at the beginning of the survey that they recognized the disease mentioned in their assigned clinical scenario. Therefore, these participants did not have to do an internet search but were immediately forwarded to the final question in which they had to indicate what they thought was the accurate diagnosis (see Figure 1). The experimental test was run from February 25 until March 12, 2017. The participants did not receive any rewards or payment for their participation but were promised to be informed about the results of the study.

Figure 1. A flowchart of patient recruitment.



Study Procedures

Participants were randomly and equally assigned to 1 of 4 clinical scenarios, which represented the symptoms of xanthelasma, seborrheic keratosis, carpal tunnel syndrome (CTS) or benign paroxysmal positional vertigo (BPPV). The scenarios of xanthelasma and seborrheic keratosis consisted of an image, whereas the scenarios of CTS and BPPV were represented through a textual description of the symptoms (see Multimedia Appendix 2). The scenarios were developed with input from clinical coauthors (NHC and JAHE). The reason why these diseases were chosen is because they belong to the so-called *minor ailments*. This term is used for relatively harmless diseases that are incredibly common in the general population and thus have a significant impact on health care. However, some of the symptoms can also occur in other diseases. As a result, they are not immediately recognizable for a medically untrained individual. Participants were not informed that the symptoms were suggestive of xanthelasma, seborrheic keratosis, CTS, or BPPV. The remaining 155 participants had to fill out a second survey at home concerning the assigned scenario and related questions (see Multimedia Appendix 3). After receiving their scenario, the participants had to choose their initial possible diagnosis that would explain the symptoms mentioned in their scenario. They also had to indicate whether they estimated that the disease was severe or not severe. The participants were then instructed to search the internet using the Web browser, Google, as though they were experiencing these symptoms themselves. Their internet search was recorded by means of print screens and search terms, which the participants recorded themselves. After conducting the internet search, they had to choose a final explanatory diagnosis, severity score, and answer questions such as on which websites they found the information to

diagnose the symptoms and whether they would discuss the information found with their physician.

Data Preparation and Coding of Education Level, Internet Search Behaviors, and Diagnostic Accuracy

Education level was classified as 1 of 3 categories: low, intermediate, and high. Participants who received primary school, lower vocational education, preparatory secondary vocational education, or general secondary education were classified as low educated. The second group who received senior secondary vocational education, senior secondary general education, or preuniversity education was classified as intermediate educated. The ones who received higher professional education or academic higher education were considered highly educated. Each search term that was entered was classified as 1 of 3 search strategies: (1) hypothesis testing, (2) narrowing within the general hypothesis area, and (3) symptom exploration [19,21]. Hypothesis testing means entering relevant search terms to verify a diagnostic primary hypothesis (ie, entering *melanoma*). The second strategy describes narrowing the search within a general hypothesis area (ie, entering *skin conditions*). Symptom exploration refers to entering search terms that involve symptoms (ie, entering *brown, hump, crust on skin*). The participants' assessments of the diagnosis were coded in 2 categories. These categories were as follows: *accurate* if the stated diagnosis matched the diagnosis of the assigned scenario and *not accurate* if the stated diagnosis did not match or if the participants did not know the answer, were unsure, or guessed multiple diagnoses. Of course, participants were not accounted for spelling errors and both medical and vernacular names were labeled as accurate, provided they were referring to the right condition. Examples of accurate diagnoses that were mentioned by participants who were assigned to the seborrheic keratosis vignette were as follows: *Senile wart* or *seborrheic keratosis*. Examples of inaccurate diagnoses were *birthmark*, *melanoma*, or *don't know*. To check whether coding was done unequivocally, almost half of the participants were also coded by a second independent team member. Team members met regularly to compare their own independent coding of participants' search terms and assessments of the diagnosis. The initial assessment of the search terms and diagnosis showed an average difference of 8.1% among team members, but this was resolved completely through discussion until members of the team reached a consensus. The quality of the websites that were used by the participants to find their diagnosis was determined by using the DISCERN instrument [24,25]. Websites were considered to be of low quality if <2, intermediate if between 2.1 and 3.9, and high if >4. Participants could submit up to four websites but were assigned to the group *low*, *intermediate*, or *high* based on the website used with the highest quality. To ensure reliable coding, a selection of 20 websites was independently classified by a second team member. All websites were independently placed in the same group, so no difference in assessment was found.

Statistical Analysis

The demographic characteristics of the participants were identified with descriptive statistics. A comparison between participants of the 4 different scenarios was made with a one-way analysis of variance (ANOVA) test (for age in years) and chi-square tests (for gender, education level, and self-estimated severity score). A McNemar test was used to compare the diagnostic accuracy of the study population before and after the internet search. Furthermore, a multivariate binary logistic regression model was constructed to predict the choice of an accurate diagnosis after searching on the Web for health information, using search strategy, age, gender, education level, clinical scenario, and the quality of the websites used as predictors. In this analysis, finding an accurate diagnosis served as the dependent variable, with finding the inaccurate diagnosis as the reference group. Finally, a McNemar test was used to compare the self-chosen severity score before and after the internet search. A *P* value lower than .05 was considered significant. All statistical analyses were performed using SPSS version 20 (IBM).

Ethical Approval

The research plan has been submitted to the Medical Ethical Committee (MEC) of the Leiden University Medical Centre. As the data cannot be traced back to the individual participant, the MEC considered the study exempt.

Results

Participant Characteristics

The participant demographic and personal characteristics are presented in Table 1. Participant demographic characteristics were identified with descriptive statistics, and a comparison between participants of the 4 different scenarios was made with a one-way ANOVA test (for age in years) and chi-square tests (for gender, education level, and self-estimated severity score). The study population ranged from 18 to 74 years of age, the overall mean age being 47.5 (SD 13.5) years. There were as many men as women (48% versus 52%, respectively). Most participants were highly educated, considering 65% graduated on a high education level and 28% obtained an intermediate education level. Comparing the 4 different scenarios with each other, the mean age was similar between the 4 groups (*P*=.45). The gender distribution is not exactly equally divided for each scenario. Especially, the scenarios of seborrheic keratosis (70% male), CTS (30% male), and BPPV (36% male) were unequally distributed (*P*=.001). Education level showed a similar distribution pattern for all 4 scenarios (*P*=.52). There was a difference in the self-estimated severity score before the internet search between the 4 different scenarios; particularly, the participants who were assigned to the seborrheic keratosis scenario tend to score the symptoms as *severe* (61%, *P*<.001).

Table 1. Participant demographic characteristics (N=155).

Characteristics per scenario	Xanthelasma (N=38)	Seborrheic keratosis (N=44)	CTS[a] (N=37)	BPPV[b] (N=36)	Total (N=155)	*P* value
Age (years), mean (SD)	50.4 (12.5)	47.1 (14.5)	47.2 (14.0)	45.3 (12.7)	47.5 (13.5)	.45
Gender, n (%)						**.001**
Male	19 (50)	31 (70)	11 (30)	13 (36)	74 (48)	
Female	19 (50)	13 (30)	26 (70)	23 (64)	81 (52)	
Education level, n (%)						**.52**
Low	2(5)	6 (14)	1 (3)	2 (6)	11 (7)	
Intermediate	12 (32)	12 (27)	11 (30)	8 (22)	43 (28)	
High	24 (63)	26 (59)	25 (68)	26 (72)	101 (65)	
Self-estimated severity score, n (%)						**<.001**
Severe	2 (5)	27 (61)	11 (30)	12 (33)	52 (34)	
Not severe	36 (95)	17 (39)	26 (70)	24 (67)	103 (66)	

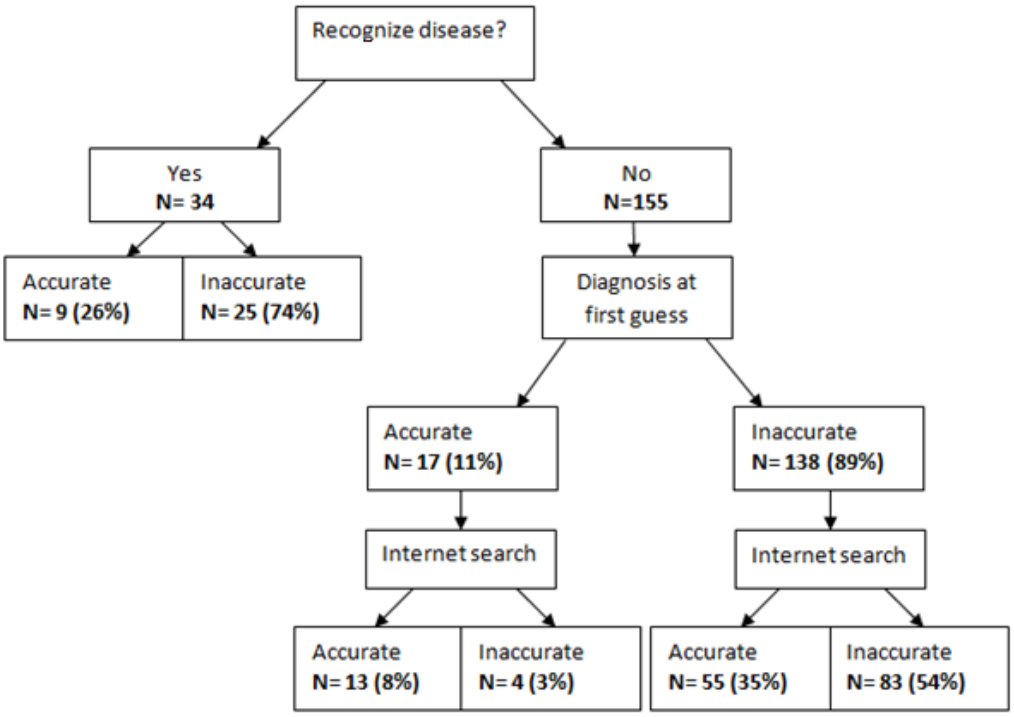
[a]CTS: carpal tunnel syndrome.

[b]BPPV: benign paroxysmal positional vertigo.

Effect of Internet Searching on Diagnostic Accuracy

Of the 189 participants, 34 indicated that they recognized the condition that caused the symptoms of the scenario, and they provided a final diagnosis without doing an internet search; only 9 out of 34 (26%) were accurate (see Figure 2). The remaining 155 participants also had to provide a first diagnosis before searching for information on the internet, of which 17 (10.9%) were accurate. After the internet search, 4 participants (4/155, 2.5%) found an incorrect diagnosis, even though their initial diagnosis was correct. Of the 138 participants who were initially inaccurate, 35.4% (55/155) found the right diagnosis. A McNemar test revealed that performing an internet search leads to a statistically significant improvement in self-diagnosing accuracy compared with before the internet search (44% versus 11%, respectively; $P<.001$).

Figure 2. The effect of internet use on self-diagnosing accuracy (N=189).



Search Strategies Used

Overall, most of the participants used the symptom exploration strategy to find a diagnosis (66%). Hypothesis testing was the second most frequently used strategy (23%), whereas the least used strategy was narrowing in the general area (12%). The used search strategy depends on the assigned scenario. A chi-square test confirms that there is a statistically significant correlation between the assigned scenario and the chosen search strategy (P<.001). The scenarios xanthelasma, CTS, and BPPV were predominantly solved using the symptom exploration strategy (68%, 92%, and 92%, respectively). Hypothesis testing was the most used strategy among the participants who were assigned to the seborrheic keratosis scenario (61%). In addition, there was a difference in the self-estimated severity score, confirmed by a chi-square test that reveals that the chosen search strategy is significantly related to the self-estimated severity score of the symptoms before the internet search (P=.001). Of the participants who used the hypothesis testing strategy, 60% scored the symptoms as *severe*, in advance. The participants who used the narrowing strategy and symptom exploration strategy only scored the symptoms as *severe* in 22% and 26% of the cases, respectively.

Characteristics That Influence Diagnostic Accuracy

The mean age of the participants who were accurate in their final diagnosis was 45.3 (SD 13.0) years. Of the participants who were inaccurate, this was 49.2 (SD 13.7; see Multimedia Appendix 4). Of the females, 52% chose the correct diagnosis, compared with 35% of the men. Furthermore, especially those participants who were assigned to diagnose xanthelasma and CTS were accurate in their diagnosis (66% and 68%, respectively). Seborrheic keratosis was the most difficult to diagnose, as only 20% chose an accurate diagnosis. In addition, users of the hypothesis testing strategy were accurate in 31% of the cases, whereas narrowing in the general area strategy led to 33% correct answers. The symptom exploration strategy led to an accurate diagnosis more often, as 50% of the individuals who used this strategy found the accurate diagnosis. Of the 196 visited websites, 6 could not be traced back by the assessor (JK). Therefore, 190 websites were checked on quality using the DISCERN instrument. Almost all visited websites were written in Dutch (95%), the remaining 5% were written in English. Of the participants who used only low-quality websites or intermediate-quality websites, 52% and 34%, respectively, found the accurate diagnosis. The participants who used at least one high-quality website were more accurate, for 64% gave a correct diagnosis. A multivariate binary logistic regression model was constructed to predict the diagnostic accuracy that was based on information retrieved from the internet. No multicollinearity was found. The resulting model showed a significant association between diagnostic accuracy with age, clinical scenario, and the quality of the websites used (see Table 2). For every 1-year increase in age, the odds of choosing the accurate diagnosis decreased by 6% (odds ratio [OR] 0.94, 95% CI 0.90 to 0.98). Furthermore, the clinical scenarios xanthelasma and CTS were significantly associated with choosing the right diagnosis (OR 10.73, 95% CI 3.24 to 35.54 versus OR 2.74, 95% CI 1.08 to 6.96, respectively) in contrast to the scenarios seborrheic keratosis (OR 0.22, 95% CI 0.07 to 0.71) and BPPV (OR 0.15, 95% CI 0.06 to 0.41), which are significantly associated with choosing an inaccurate diagnosis. Another factor significantly related to finding the right diagnosis is the highest quality of the websites used on which the participant found the final diagnosis. Compared with participants who only used low-quality websites, participants who used at least one high-quality website were most likely to find the accurate diagnosis (OR 7.49, 95% CI 1.85 to 30.26), followed by intermediate-quality website users (OR 1.53, 95% CI 0.33 to 7.00). In addition, there was no significant association with gender, education level, and search strategy.

Table 2. Multivariate binary logistic regression analysis of the relationship of demographic characteristics with reliance on choosing the accurate diagnosis after an internet search (N=155).

Characteristics	Odds ratio (95% CI)	*P* value
Age (years)	0.94 (0.90-0.98)	.005
Gender		**.12**
Male	1 (reference)	
Female	2.36 (0.80-6.97)	
Education level		**.82**
Low	1 (reference)	
Intermediate	0.81 (0.09-7.21)	
High	0.62 (0.09-4.31)	
Scenario[a]		**<.001**
Xanthelasma	10.73 (3.24-35.54)	
Seborrheic keratosis	0.22(0.070-0.71)	
CTS[b]	2.74 (1.08-6.96)	
BPPV[c]	0.15 (0.06-0.41)	
Search strategy[a]		**.82**
Hypothesis testing	1.41 (0.48-4.18)	
Narrowing	0.83 (0.24-2.81)	
Symptom exploration	0.86 (0.33-2.21)	
Quality of websites used		**.005**
Low	1 (reference)	
Intermediate	1.53 (0.33-7.00)	
High	7.49 (1.85-30.26)	

[a]The odds ratios of *Scenario* and *Search strategy* were derived by settings, CONTRAST subcommand deviation. The effect for each category of the independent variable is compared with the overall mean.

[b]CTS: carpal tunnel syndrome.

[c]BPPV: benign paroxysmal positional vertigo.

Influence of Searching the Internet on Self-Estimated Severity Score

Before searching the internet, 34% of the participants tend to score the symptoms of the assigned scenario as *severe*. After finding information about the possible diagnosis on the internet, 34% still scored the symptoms as *severe*. Therefore, searching the internet did not lead to a statistically significant change in participants' beliefs about the severity of the condition (McNemar test, *P*=.85). This last self-chosen severity score after the internet search was related to the assigned scenario (chi-square test, *P*=.002) but was not related to the diagnostic accuracy (chi-square test, *P*=.13).

Whether or Not to Discuss With the Physician?

In general, almost two-thirds (65%) of the participants have discussed medical information found on the Web with their physician in the past. In 75% of these consultations, the participants received a positive response from the physician. The most frequently heard comments from the physicians to their patients were that searching the internet contributes to well-informed patients and that it makes them better prepared for the consultation. Furthermore, it was mentioned that the physicians need to explain less and therefore the consultation would take less time. Only 5% received a negative response, which mainly consisted of warnings that the information on the internet is not always correct and that self-diagnosing by laymen is undesirable. Of the other participants who did not share the information found on the Web with their physician, most did not have a special reason for not telling or had simply forgotten to do so. Only 3 participants mentioned that they expected that, after sharing, the physician would no longer look at them with an *open mind*.

Discussion

Principal Findings and Comparison With Prior Work

With our vignette-based experimental study, we tried to provide an initial insight into how patients search for medical information on the internet and how they attempt to self-diagnose symptoms of minor ailments on the Web. Despite the fact that Web-based self-diagnosis is very popular, limited research has been done on this topic, especially regarding

internet search on minor ailments, even though these ailments are common health issues and therefore a very common reason for consulting the GP. Our findings suggest that searching on the Web can be a helpful tool in the process of self-diagnosing. Overall, 44% of the participants were accurate in their final diagnosis, which was a significant improvement compared with the 11% who were accurate before searching the internet. On the contrary, it should also be considered that 3% of the participants who were initially accurate found an incorrect diagnosis afterward. Of the participants who stated they recognize the diagnosis in advance without the help of the internet, the self-diagnosing accuracy was poor and only 26% were accurate.

To determine why some participants find the right diagnosis whereas others do not, we observed how people search for medical information on the internet and which factors or characteristics contribute to finding the accurate diagnosis. In the overall study group, most people tend to use a symptom exploration strategy. Hypothesis verification strategy and narrowing in the general area strategy were used less. This corresponds with previous research on internet search behaviors of emergency department patients and a study on internet queries, in which it is also shown that most of the internet searches are focused on symptoms [17,26]. Which strategy was chosen seems to be associated with the type of scenario and how severe the participants estimated that the diagnosis would be in advance. For example, a lot of participants assigned to the seborrheic keratosis scenario filled in an initial diagnosis of melanoma or skin cancer and estimated that the diagnosis would be severe. A significant majority of the *keratosis* participants chose the hypothesis testing strategy. This is in line with other previous research that showed that more focused seekers usually have a clear idea and a plan to research in a limited set of results (hypothesis testing), whereas more exploratory seekers usually try to address unfamiliar problems by retrieving a wider range of information (symptom exploration). Patients who are searching for information about medical issues tend to use a more exploratory method, especially individuals who need to handle health issues of themselves or their loved ones and individuals who have a high level of uncertainty about the subject [27]. After the internet search for the seborrheic keratosis scenario with the hypothesis testing strategy, many participants still thought the diagnosis would be skin cancer, and this scenario had the most inaccurate participants compared with the other three. This phenomenon was observed more often in previous research and is known as confirmation bias (starting with a hypothesis and only looking at information that confirms the initial hypothesis) and premature termination bias (quitting the search after viewing only 1 topic) [19,28]. These studies show that especially hypothesis-driven strategies are prone to these forms of bias.

Furthermore, the multivariate binary logistic regression analysis showed that younger age, the symptom scenario xanthelasma and CTS, and the use of higher-quality websites lead to more self-diagnosing accuracy. In the current literature, younger age and higher education are characteristics that contribute to search strategies that are more successful in finding an accurate diagnosis [21]. Our study confirms that a younger age is significantly related to diagnostic accuracy. A possible explanation for this is that younger people have more experience with browsing the internet and that this contributes to finding the right diagnosis [20]. However, the literature is contradictory, as another study stated that more internet experience was not significantly related to diagnostic accuracy [19]. Our study did not find a significant relation with higher education but that may be caused by the fact that few lower educated participants were included in our study population (see Limitations section). It could be based on coincidence that our lower educated individuals just did well self-diagnosing, and therefore we could not detect a significant difference compared with the higher educated participants. A possible explanation for why the scenarios xanthelasma and CTS were easier for the participants to diagnose is because the differential diagnosis of these symptoms is less extensive and therefore possibly easier to find on the internet. Seborrheic keratosis was the most difficult to diagnose, as only 20% chose an accurate diagnosis. Despite the unequal distribution in gender in this scenario, no difference was found between men (19% accurate) and women (23% accurate). Furthermore, we found that participants who used higher-quality websites were more likely to choose the accurate diagnosis. After studying the list of websites used, it can be concluded that the websites that are considered to be of high quality according to the DISCERN method have often been developed by hospitals or doctors' associations or departments. Examples are thuisarts.nl (Dutch College of General Practitioners) or oogartsen.nl (Dutch foundation for ophthalmologists of top clinical hospitals). The websites classified as low quality are mainly health forums, where everyone can write anything and there is little or no control over the quality of this information, for example, Artikelsite.info or Mens-en-gezondheid.infonu.nl. This demonstrates that it is of great importance to guide lay people in their internet search by offering websites with reliable information that has been written or verified by professionals. This concept corresponds to previous research that examined the possible barriers and needs of patients searching for medical information on the internet [29]. One important finding was that patients have indicated to prefer guidance of health professionals to find appropriate Web-based resources [29,30]. Furthermore, the findings suggest that greater involvement by health professionals could contribute to an improved relationship between the health professional and patient by minimizing barriers such as finding incomprehensible information, information in medical jargon, volume of information, and inconsistency of information across different sources [29]. We suggest that health professionals can play a role in consumers' navigation of Web-based health information by, for example, placing weblinks to high-quality websites on the internet homepage of the doctor's practice or on the information board that is often present in the waiting room of the practice.

What falls outside the scope of our study is what role search engines play in managing the displayed content and what impact this has on the decision-making process of the consumer. People often only view the top search results and what is shown there is logically very important for information gathering. Some recent researches show that the bias of search engines or social media can have an effect on people [31,32]. There is evidence

that obscuring the true identity of an information source, obscuring the affiliations of an information source, and control over user-generated content can greatly influence consumer health knowledge and behavior [31]. It is logical to think that this can also influence the self-diagnosing process, but how exactly is an important subject for future research.

In addition, a frequently mentioned reason why people search for Web-based medical information is to find reassurance [2]. However, the results of the survey show that the self-estimated severity score in advance does not differ from the self-estimated severity score after searching the internet, and this is not related to whether the participants find the right diagnosis or not. Therefore, our study shows that people actually do not find the reassurance they are looking for, even if they find the right diagnosis. Apparently, a visit to the doctor is still necessary. Almost two-thirds of the participants indicated that they would share the information they found with their physician. Previous research supports these findings and indicates that the most frequently mentioned reason to discuss the information found with their physician is to ascertain the opinions of health professionals on the retrieved health information [30]. Most participants received a positive response from their doctor.

Limitations

There are factors that limit the generalization of the findings. First, the surveys were conducted among panel members of the Dutch Consumers' Association. This is an independent nonprofit association that conducts research and makes publications about various products, services, and injustice in society. Anyone can become a member of this organization; however, one can expect that these individuals are more conscious and are higher educated compared with the rest of the Dutch population. By selecting these subjects, in particular, a form of selection bias occurs. That would explain why our study population is so highly educated. As some studies show that finding the right diagnosis is related to a higher level of education, it is expected that finding the right diagnosis in the general population is even more difficult than finding the right diagnosis in our research group. Second, participants were instructed to do a scenario-based internet search and were therefore driven by symptoms they did not experience at that moment. Furthermore, in reality it is possible that the course of a disease is gradual and has different stages (eg, CTS). Therefore, patients may experience different symptoms over time instead of perceiving all the symptoms at the same time. This can make the search for the correct diagnosis more difficult in real life. These 2 factors might influence the generalizability of our results, as it may have artificially influenced participants' search efforts and strategies. Finally, it should be considered that the chosen search strategy and whether patients find the right diagnosis or not depends on the disease and experienced symptoms. Therefore, we realize that only these 4 chosen scenarios cannot represent how patients search the internet for medical information in general, but it does provide an indication.

Conclusions

Our findings suggest that most patients who search the internet for medical information use a symptom-based approach, but this depends on the experienced symptoms. If the patient expects the potential diagnosis to be severe, they tend to use a hypothesis verification strategy more often and are therefore prone to certain forms of bias. To prevent this, doctors should advise their patients to look for symptoms rather than hypothesis-driven strategies. In addition, self-diagnosing accuracy is related to younger age, the symptom scenario, and the use of higher-quality websites. Although it is difficult to tackle the abundance of low-quality websites, doctors should focus on ways to guide patients toward professional websites that may be more likely to lead to accurate decision making. This can be archived, for example, by placing weblinks to high-quality websites on the internet homepage of the doctor's practice or on the information board that is often present in the waiting room of the practice. However, future research that includes more patients and more different types of scenarios will be necessary to further understand the complex coordination between patient search strategies, finding reliable websites, and Web-based symptom information processing.

Acknowledgments

The authors would like to thank Dr ir Nan van Geloven for her review of subsequent versions of this paper. Furthermore, the authors would like to acknowledge Irene AEI de Best for her feedback with regard to the grammar and readability, making this paper as it is now.

Conflicts of Interest

None declared.

Multimedia Appendix 1
First survey Dutch Consumers' Association.
[PDF File (Adobe PDF File), 52 KB - jmir_v21i11e12278_app1.pdf]

Multimedia Appendix 2
Clinical scenarios.
[PDF File (Adobe PDF File), 392 KB - jmir_v21i11e12278_app2.pdf]

Multimedia Appendix 3
Second survey internet search.
[PDF File (Adobe PDF File), 42 KB - jmir_v21i11e12278_app3.pdf]

Multimedia Appendix 4
Table with characteristics by accuracy of diagnosis.
[PDF File (Adobe PDF File), 39 KB - jmir_v21i11e12278_app4.pdf]

References

1. Wimble M. Understanding health and health-related behavior of users of internet health information. Telemed J E Health 2016 Oct;22(10):809-815. [doi: 10.1089/tmj.2015.0267] [Medline: 27045569]
2. Mokkink DH, Dijkstra R, Verbakel D. Huisarts & Wetenschap. 2008. Surfende Patiënten URL: https://www.henw.org/artikelen/surfende-patienten [accessed 2018-08-23] [WebCite Cache ID 71t1PI0qb]
3. Teriaky A, Tangri V, Chande N. Use of internet resources by patients awaiting gastroenterology consultation. Turk J Gastroenterol 2015 Jan;26(1):49-52 [FREE Full text] [doi: 10.5152/tjg.2015.6043] [Medline: 25698271]
4. Rice RE. Influences, usage, and outcomes of internet health information searching: multivariate results from the Pew surveys. Int J Med Inform 2006 Jan;75(1):8-28. [doi: 10.1016/j.ijmedinf.2005.07.032] [Medline: 16125453]
5. Brewer NT. The impact of internet use on health cognitions and behaviour. J Consum Psychol 2005;433(453):405-424 [FREE Full text]
6. Mueller J, Jay C, Harper S, Davies A, Vega J, Todd C. Web use for symptom appraisal of physical health conditions: a systematic review. J Med Internet Res 2017 Jun 13;19(6):e202 [FREE Full text] [doi: 10.2196/jmir.6755] [Medline: 28611017]
7. Spoelman WA, Bonten TN, de Waal MW, Drenthen T, Smeele IJ, Nielen MM, et al. Effect of an evidence-based website on healthcare usage: an interrupted time-series study. BMJ Open 2016 Nov 9;6(11):e013166 [FREE Full text] [doi: 10.1136/bmjopen-2016-013166] [Medline: 28186945]
8. Oh YS, Song NK. Investigating relationships between health-related problems and online health information seeking. Comput Inform Nurs 2017 Jan;35(1):29-35. [doi: 10.1097/CIN.0000000000000234] [Medline: 26950091]
9. Joury AU, Alshathri M, Alkhunaizi M, Jaleesah N, Pines JM. Internet websites for chest pain symptoms demonstrate highly variable content and quality. Acad Emerg Med 2016 Oct;23(10):1146-1152 [FREE Full text] [doi: 10.1111/acem.13039] [Medline: 27341392]
10. Hirsch M, Aggarwal S, Barker C, Davis CJ, Duffy JM. Googling endometriosis: a systematic review of information available on the internet. Am J Obstet Gynecol 2017 May;216(5):451-8.e1. [doi: 10.1016/j.ajog.2016.11.1007] [Medline: 27840143]
11. Cardel MI, Chavez S, Bian J, Peñaranda E, Miller DR, Huo T, et al. Accuracy of weight loss information in Spanish search engine results on the internet. Obesity (Silver Spring) 2016 Nov;24(11):2422-2434 [FREE Full text] [doi: 10.1002/oby.21646] [Medline: 27653438]
12. Newnham GM, Burns WI, Snyder RD, Dowling AJ, Ranieri NF, Gray EL, et al. Information from the internet: attitudes of Australian oncology patients. Intern Med J 2006 Nov;36(11):718-723. [doi: 10.1111/j.1445-5994.2006.01212.x] [Medline: 17040358]
13. Diaz JA, Griffith RA, Ng JJ, Reinert SE, Friedmann PD, Moulton AW. Patients' use of the internet for medical information. J Gen Intern Med 2002 Mar;17(3):180-185 [FREE Full text] [doi: 10.1046/j.1525-1497.2002.10603.x] [Medline: 11929503]
14. Powell J, Inglis N, Ronnie J, Large S. The characteristics and motivations of online health information seekers: cross-sectional survey and qualitative interview study. J Med Internet Res 2011 Feb 23;13(1):e20 [FREE Full text] [doi: 10.2196/jmir.1600] [Medline: 21345783]
15. Dwyer DS, Liu H. The impact of consumer health information on the demand for health services. Q Rev Econ Finance 2013 Feb;53(1):1-11. [doi: 10.1016/j.qref.2012.11.002]
16. Suziedelyte A. How does searching for health information on the internet affect individuals' demand for health care services? Soc Sci Med 2012 Nov;75(10):1828-1835. [doi: 10.1016/j.socscimed.2012.07.022] [Medline: 22884947]
17. McCarthy DM, Scott GN, Courtney DM, Czerniak A, Aldeen AZ, Gravenor S, et al. What did you google? Describing online health information search patterns of ED patients and their relationship with final diagnoses. West J Emerg Med 2017 Aug;18(5):928-936 [FREE Full text] [doi: 10.5811/westjem.2017.5.34108] [Medline: 28874946]
18. Pang PC, Chang S, Verspoor K, Pearce J. Designing health websites based on users' web-based information-seeking behaviors: a mixed-method observational study. J Med Internet Res 2016 Jun 6;18(6):e145 [FREE Full text] [doi: 10.2196/jmir.5661] [Medline: 27267955]
19. Keselman A, Browne AC, Kaufman DR. Consumer health information seeking as hypothesis testing. J Am Med Inform Assoc 2008;15(4):484-495 [FREE Full text] [doi: 10.1197/jamia.M2449] [Medline: 18436912]
20. Luger TM, Houston TK, Suls J. Older adult experience of online diagnosis: results from a scenario-based think-aloud protocol. J Med Internet Res 2014 Jan 16;16(1):e16 [FREE Full text] [doi: 10.2196/jmir.2924] [Medline: 24434479]

21. Perez SL, Paterniti DA, Wilson M, Bell RA, Chan MS, Villareal CC, et al. Characterizing the processes for navigating internet health information using real-time observations: a mixed-methods approach. J Med Internet Res 2015 Jul 20;17(7):e173 [FREE Full text] [doi: 10.2196/jmir.3945] [Medline: 26194787]
22. Perez SL, Kravitz RL, Bell RA, Chan MS, Paterniti DA. Characterizing internet health information seeking strategies by socioeconomic status: a mixed methods approach. BMC Med Inform Decis Mak 2016 Aug 9;16:107 [FREE Full text] [doi: 10.1186/s12911-016-0344-x] [Medline: 27506607]
23. Fiksdal AS, Kumbamu A, Jadhav AS, Cocos C, Nelsen LA, Pathak J, et al. Evaluating the process of online health information searching: a qualitative approach to exploring consumer perspectives. J Med Internet Res 2014 Oct 7;16(10):e224 [FREE Full text] [doi: 10.2196/jmir.3341] [Medline: 25348028]
24. Charnock D, Shepperd S, Needham G, Gann R. DISCERN: an instrument for judging the quality of written consumer health information on treatment choices. J Epidemiol Community Health 1999 Feb;53(2):105-111 [FREE Full text] [doi: 10.1136/jech.53.2.105] [Medline: 10396471]
25. Charnock D, Shepperd S, Needham G, Gann R. DISCERN. The DISCERN Instrument URL: http://www.discern.org.uk/discern_instrument.php [accessed 2018-08-31] [WebCite Cache ID 725M6hc7a]
26. Jadhav A, Andrews D, Fiksdal A, Kumbamu A, McCormick JB, Misitano A, et al. Comparative analysis of online health queries originating from personal computers and smart devices on a consumer health information portal. J Med Internet Res 2014 Jul 4;16(7):e160 [FREE Full text] [doi: 10.2196/jmir.3186] [Medline: 25000537]
27. Pang PC, Chang S, Pearce J, Verspoor K. Online Health Information Seeking Behaviour: Understanding Different Search Approaches. In: Proceedings of the Pacific Asia Conference on Information Systems. 2014 Presented at: PACIS'14; June 24-28, 2014; Chengdu, China.
28. Mueller J, Jay C, Harper S, Todd C. The role of web-based health information in help-seeking behavior prior to a diagnosis of lung cancer: a mixed-methods study. J Med Internet Res 2017 Jun 8;19(6):e189 [FREE Full text] [doi: 10.2196/jmir.6336] [Medline: 28596146]
29. Lee K, Hoti K, Hughes JD, Emmerton L. Dr Google and the consumer: a qualitative study exploring the navigational needs and online health information-seeking behaviors of consumers with chronic health conditions. J Med Internet Res 2014 Dec 2;16(12):e262 [FREE Full text] [doi: 10.2196/jmir.3706] [Medline: 25470306]
30. Lee K, Hoti K, Hughes JD, Emmerton LM. Consumer use of 'Dr Google': a survey on health information-seeking behaviors and navigational needs. J Med Internet Res 2015 Dec 29;17(12):e288 [FREE Full text] [doi: 10.2196/jmir.4345] [Medline: 26715363]
31. Kulshrestha J, Eslami M, Zafar MB, Ghosh S, Gummadi KP, Karahalios K. Quantifying Search Bias: Investigating Sources of Bias for Political Searches in Social Media. In: Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing. 2017 Presented at: CSCW'17; February 25-March 1, 2017; Portland, Oregon, USA. [doi: 10.1145/2998181.2998321]
32. DeAndrea DC, Vendemia MA. How affiliation disclosure and control over user-generated comments affects consumer health knowledge and behavior: a randomized controlled experiment of pharmaceutical direct-to-consumer advertising on social media. J Med Internet Res 2016 Jul 19;18(7):e189 [FREE Full text] [doi: 10.2196/jmir.5972] [Medline: 27435883]

Abbreviations

ANOVA: analysis of variance
BPPV: benign paroxysmal positional vertigo
CTS: carpal tunnel syndrome
GP: general practitioner
MEC: Medical Ethical Committee
OR: odds ratio

Edited by G Eysenbach; submitted 25.09.18; peer-reviewed by C Jay, L Wang, J Suls, J Bian, PCI Pang, A Bruton; comments to author 29.11.18; revised version received 22.01.19; accepted 04.03.19; published 11.11.19.

Please cite as:
Kwakernaak J, Eekhof JAH, De Waal MWM, Barenbrug EAM, Chavannes NH
Patients' Use of the Internet to Find Reliable Medical Information About Minor Ailments: Vignette-Based Experimental Study
J Med Internet Res 2019;21(11):e12278
URL: http://www.jmir.org/2019/11/e12278/
doi:10.2196/12278
PMID:31710304

Original Paper

Electronic Health Literacy and Dietary Behaviors in Taiwanese College Students: Cross-Sectional Study

Shu Ching Yang[1*], PhD; Yi Fang Luo[1*], PhD; Chia-Hsun Chiang[1*], PhD

Institute of Education, National Sun Yat-Sen University, Kaohsiung, Province of China Taiwan

[*]all authors contributed equally

Corresponding Author:
Chia-Hsun Chiang, PhD
Institute of Education
National Sun Yat-Sen University
70 Lienhai Rd
Kaohsiung
Taiwan
Phone: 886 7 5251521
Email: d996050002@student.nsysu.edu.tw

Abstract

Background: Given the recognized importance of preventing poor dietary behaviors during adolescence, we need a better understanding of college students' dietary behaviors. Studies have found that individual factors and electronic health (eHealth) literacy may affect one's dietary behaviors. However, few studies have fully investigated the effect of the three levels of eHealth literacy (functional, interactive, and critical) and the interactive effect of individual factors (eg, gender, monthly expenses, and frequency of cooking) and the three levels of eHealth literacy on the four aspects of dietary behaviors (consumer health, balanced diet, regular eating habits, and unhealthy food intake).

Objective: This study aimed to investigate whether individual differences and higher eHealth literacy are associated with more positive dietary behaviors and less unhealthy dietary intake.

Methods: The eHealth Literacy Scale is a 12-item instrument designed to measure college students' functional, interactive, and critical eHealth literacy. The Dietary Behaviors Scale is a 14-item instrument developed to measure four aspects of dietary behaviors of college students. A questionnaire was administered to collect background information about participants' gender, monthly expenses, and frequency of cooking. A national sample of college students was surveyed, and 813 responses were obtained. We conducted a multiple regression analysis to examine the association among individual factors, eHealth literacy, and dietary behaviors.

Results: This study found that functional eHealth literacy was negatively related to unhealthy food intake (beta=−.11; *P*=.01), and interactive eHealth literacy was positively related to balanced diet (beta=.25; *P*<.001) and consumer health (beta=.15; *P*=.02). Moreover, critical eHealth literacy was positively related to consumer health (beta=.30; *P*<.001) and regular eating habits (beta=.20; *P*=.002). Finally, the interactive effect between gender and interactive eHealth literacy was negatively related to balanced diet (beta=−.22; *P*<.001). The interactive effect between monthly expenses and functional eHealth literacy was positively related to balanced diet (beta=.07; *P*=.03), although the interactive effect between monthly expenses and critical eHealth literacy was negatively related to balanced diet (beta=−.10; *P*=.047).

Conclusions: This study showed that Taiwanese college students with higher functional eHealth literacy were more likely to engage in fewer unhealthy food consumption practices. Those who had higher interactive and critical eHealth literacy were more likely to engage in positive dietary behaviors than those with functional eHealth literacy. Surprisingly, females with high interactive eHealth literacy were more likely to have a poor balanced diet. In contrast, students with higher monthly expenses and higher functional eHealth literacy were more likely to have a balanced diet. However, students with higher monthly expenses and higher critical eHealth literacy were less likely to maintain a balanced diet.

(J Med Internet Res 2019;21(11):e13140) doi:10.2196/13140

KEYWORDS

college; dietary; health literacy; students

Introduction

Background

The importance of preventing poor dietary behaviors during adolescence has been recognized because of its impact on long-term health issues such as the development of obesity and other noncommunicable diseases [1]. Students' dietary behaviors are most affected by their families and schools before they start college, but afterward, their dietary choices are mostly made independently [2]. The college years offer an opportunity for new experiences and personal freedom, such as the autonomy of food choices; however, the period is also noted for the emergence of unhealthy behaviors that place college students at risk of health problems [3]. Noteworthy findings about college students' unhealthy eating behaviors reveal frequent consumption of snacks and fried foods and of fewer than 5 servings of fruits and vegetables daily [4,5]. Moreover, college students may put themselves at increased risk by consuming unsafe foods that may cause foodborne illness (eg, undercooked meat and shellfish containing marine toxins) or by not following accepted food safety practices [6]. Clearly, campus life should provide college students with crucial opportunities to develop healthy eating behaviors and adopt a nutritious and balanced food intake, thereby forming a strong basis for good health throughout life. Thus, we need a better understanding of college students' dietary behaviors.

Dietary behaviors, including consumer health, balanced diet, regular eating habits, and unhealthy food intake, can be viewed as positive or negative actions in relation to maintaining or enhancing health [7]. Consumer health refers to individuals' health awareness and decisions about food purchasing. A balanced diet indicates eating the recommended amount of a variety of foods from each food group daily. Regular eating habits refer to individuals' healthy eating habits. Unhealthy food intake refers to individuals' consumption of junk food and drinks [7]. Individuals must understand the prevailing nutritional recommendations with respect to the food products that they are considering and use such recommendations to choose suitable foods [8]. These practices and skills are included in the concept of health literacy.

Health literacy refers to an individual's ability to access, understand, and use information to promote and maintain good health [9]. Studies have shown that individuals with higher health literacy are more likely to use food labels and have higher dietary quality [10], healthier eating behavior, and lower consumption of sugar-sweetened beverages [11]. Recently, the advent of the internet has drastically changed how health information is disseminated, and the internet is now widely used to obtain this information. People should not only be health literate but also have the capabilities, resources, and motivation to find, understand, and appraise health information using digital services and technology [12]. Unlike other distinct forms of literacy, electronic health (eHealth) literacy combines basic literacy as well as information, health, media, computer, and scientific literacies and applies them to eHealth promotion [13]. To obtain a complete overview of people's skills in obtaining and using health information, it is more necessary to measure eHealth literacy than to measure health literacy [14].

Norman and Skinner [13] have defined eHealth literacy, which includes functional, interactive, and critical levels [15], as "the ability to seek, find, understand, and appraise health information from electronic sources and to apply the knowledge gained to addressing or solving health problems.". Functional eHealth literacy refers to basic skills in reading and writing about Web-based health information. Interactive eHealth literacy specifies the communicative and social skills that can be used to extract information in social Web-based multimedia environments. Critical eHealth literacy involves the cognitive skills that can be applied to critically evaluate the credibility, relevance, and risks of sharing and receiving Web-based health information [16-18]. Such a classification indicates that the different levels of literacy progressively allow for greater autonomy and personal empowerment in decision making as well as engagement in a wider range of health actions [9,15]. Having the composite skills of eHealth literacy allows individuals to achieve positive health behaviors [19].

Objectives

The integrative model of eHealth use (IMeHU) indicates that people with high eHealth literacy are not only more inclined to use the internet to find answers to health-related questions but are also able to understand the information that they find, verify its veracity, and use it to promote health behaviors [20]. Studies have found that 27 % web user reported that their eating habits had affected by eHealth resources (eg, health websites) [21]. In addition, individuals with higher eHealth literacy adopt more balanced nutrition [22]. Recently, researchers have focused on the three levels of eHealth literacy and found that individuals with high functional [16] and critical eHealth literacy [16,17] are more likely to practice balanced eating habits. However, other studies have found no statistically significant evidence of a relationship between interactive eHealth literacy and eating habits [16,17]. A systematic review demonstrated few studies reporting an association between eHealth literacy and health behaviors, with inconsistent results [19]. Accordingly, this study attempted to fully investigate the roles of functional, interactive, and critical eHealth literacy on the four aspects of dietary behaviors. We thus propose the following hypothesis: H1: college students who possess higher functional, interactive, and critical eHealth literacy will engage in more positive dietary behaviors.

The IMeHU suggests that individual factors may affect one's health behaviors [20]. Studies have found that college students with higher meal expenses were more likely to display poor dietary behaviors [23,24]. Cooking skills are important for health and well-being [25], as studies have shown that an intervention involving gardening, nutrition, and cooking can improve dietary intake [26]. Some studies have also reflected gender differences, such as 1 study that showed girls' higher engagement in dietary behaviors than boys [27]. However, other studies have revealed female college students' insufficient vegetable intake [28] and a poor balanced diet and unhealthier food intake than that of male students [7]. Given the abovementioned studies on individual differences in dietary behaviors, this study performed

analyses of individual factors (eg, gender, monthly expenses, and frequency of cooking) and the three levels of eHealth literacy to examine the explanatory power of the four aspects of dietary behaviors.

Similarly, little attention has been paid to the interactive effect of individual factors and the three levels of eHealth literacy on the four aspects of dietary behaviors. It is necessary to clarify the effects of individual factors and the three levels of eHealth literacy, as well as how the interactive effect of these 2 elements affects the four aspects of dietary behaviors, to develop an effective health education program to promote healthy dietary behaviors among college students. Therefore, the objective of this study was to investigate the associations among individual factors, eHealth literacy, and dietary behaviors in Taiwanese college students with the goal of determining whether high eHealth literacy is associated with more positive dietary behaviors and less unhealthy dietary intake and to investigate the interactive effects of individual factors and eHealth literacy on the four aspects of dietary behaviors. In accordance with the IMeHU [20] and the abovementioned studies, we therefore propose the following hypothesis: H2: individual factors and eHealth literacy have an interactive effect on the dietary behaviors of college students.

Methods

Participants

This study was conducted nationwide in Taiwan. The list of colleges came from Taiwan's Ministry of Education. Sampling units were drawn from within each region. We contacted teachers at selected colleges to request their assistance in the pen-and-paper survey distribution. Ultimately, 1100 college students from 10 schools were recruited to participate in the survey. Of the 895 surveys received, we excluded 25 blank surveys, 34 surveys with incomplete respondents' individual factors, and 23 surveys with fixed mode answers. However, 21 surveys with fewer than 3 items incomplete were retained. We applied expectation-maximization method of multiple imputation to the missing data. In other words, 813 surveys were statistically analyzed in this study (813/895, 90.8%). We gathered the respondents' individual factors, including information about age, gender (male and female groups), monthly expenses (<NT $5000, NT $5000-NT $10,000, and >NT $10,001), and frequency of cooking (5-point Likert scale). The frequency of cooking was measured by asking how often the students cooked by themselves and was rated based on the responses on a scale from 1 (never) to 5 (always).

The Survey Instrument

Electronic Health Literacy Scale

Participants' eHealth literacy was evaluated using the eHealth Literacy Scale (eHLS) [17], which consists of 12 items that constitute 3 levels: functional (3 items), interactive (4 items), and critical (5 items) eHealth literacy. The functional level evaluates individuals' basic reading and writing skills and basic knowledge of health conditions and health systems. The interactive level evaluates individuals' communicative and social skills, which can be used to abstract information and derive meaning from different forms of communication. The critical level assesses individuals' most advanced cognitive skills, which can be applied to critically analyze information, discern the quality of health websites, and use good information to make informed decisions about health.

The items were answered using a 5-point Likert scale, with scores ranging from 1 (total disagreement) to 5 (total agreement). High scores on the respective levels indicated higher functional, interactive, and critical eHealth literacy. Cronbach alpha is a statistic commonly quoted by authors to demonstrate that tests and scales that have been constructed or adopted for research projects are fit for the research purpose. The alpha values of .45 to .98 for this instrument have been described as acceptable [29]. Within the sample used in this study, the Cronbach alpha values of functional eHealth literacy, interactive eHealth literacy, and critical eHealth literacy were .81, .87, and .90, respectively, indicating that the eHLS had acceptable internal reliability.

Dietary Behaviors Scale

The participants' dietary behaviors were evaluated using the Dietary Behaviors Scale (DBS) [7]. Using item analyses, exploratory factor analysis, and confirmatory factor analysis, Luo et al [7] revealed that the DBS is a reliable and validated measure of dietary behaviors of Taiwanese college students. The DBS consists of 14 items that comprise 4 aspects: consumer health (4 items), balanced diet (4 items), regular eating habits (3 items), and unhealthy food intake (3 items). The consumer health aspect evaluates individuals' attitudes and decisions about food purchasing. The balanced diet aspect evaluates individuals' nutrient intake. The regular eating habits aspect evaluates individuals' healthy eating habits. The unhealthy food intake aspect evaluates individuals' consumption of food and drinks that are high in calories, fat, salt, or sugar.

The items were answered using a 5-point Likert scale, with scores ranging from 1 (never) to 5 (always). High scores in the individual aspects indicated positive attitudes and decisions about product purchases, more balanced and regular eating habits, and more consumption of unhealthy food. Within the sample used in this study, the Cronbach alpha values of consumer health, balanced diet, regular eating habits, and unhealthy food intake were .82, .72, .59, and .68, respectively, indicating that the DBS had acceptable internal reliability [29].

Data Analysis

Analyses were conducted using SPSS 17.0 (IBM Corp). We performed 4 multiple regression analyses to examine the explanatory power of the four aspects of dietary behaviors. In the model, individual factors, the 3 levels of eHealth literacy, and the interaction of these 2 elements were entered. Gender was viewed as a dummy variable where males were coded as 0, and monthly expenses and frequency of cooking were viewed as continuous variables.

Ethical Considerations

The study adopted an anonymous questionnaire in line with the government's institutional review board rules for exempt review. The questionnaire instructions informed the participants of the

research purpose and confidentiality and indicated that they had the right to refuse to participate at any time. The participants received the questionnaire and gifts at the same time; even if a participant decided to drop out of the investigation, he or she still received the gifts (a pen and an L-folder). This approach was intended to be fair to each participant, to avoid the impact of gift incentives on the participants, and to provide compensation for the participants.

Results

Participants' Demographics and Characteristics

Table 1 presents the demographics and characteristics of the study participants. The mean age of participants was 20.08 (SD 1.43) years. Of the 813 participants, 47.1% (383/813) were female and 40.7% (331/813) reported that their monthly expenses were less than NT $5000. The mean frequency of cooking of the participants was 2.23 (SD 0.96).

Table 1. Demographics and characteristics of the sample.

Variables	n (%)
Gender	
Male	430 (52.9)
Female	383 (47.1)
Monthly expenses	
<NT $5000	331 (40.7)
NT $5001-NT $10,000	409 (50.3)
>NT $10,001	73 (9.0)
Frequency of cooking	
Never	180 (22.2)
Seldom	373 (45.9)
Sometimes	166 (20.4)
Often	80 (9.8)
Always	14 (1.7)

Descriptive Statistics of Electronic Health Literacy and Dietary Behaviors

Among all participants, the mean scores of functional eHealth literacy, interactive eHealth literacy, and critical eHealth literacy were 3.56 (SD 0.77), 3.57 (SD 0.71), and 3.59 (SD 0.72), respectively, indicating that college students had medium or higher levels of eHealth literacy.

On the DBS, the mean scores of the regular eating habits, balanced diet, unhealthy food intake, and consumer health were 3.15 (SD 0.73), 2.92 (SD 0.70), 3.02 (SD 0.72), and 3.12 (SD 0.76), respectively. This result indicated that college students had positive consumer health and regular eating habits although they did not maintain a balanced diet. Moreover, they sometimes ate unhealthy food.

Analysis of Electronic Health Literacy and Dietary Behaviors

In the multiple regression analysis, first, the tolerance and variance inflation factor (VIF), as 2 collinearity diagnostic factors, were examined to assess multicollinearity in this study. The collinearity diagnostics results showed that tolerance ranged from 0.28 to 0.99, and VIF ranged from 1.01 to 3.78. These results indicate that explanatory variables in this multiple regression model were weakly linearly related [30].

The results of the multiple regression analysis are displayed in Multimedia Appendix 1, which shows that functional eHealth literacy was negatively related to unhealthy food intake (beta=−.11; P=.01), and interactive eHealth literacy was positively related to balanced diet (beta=.25; P<.001) and consumer health (beta=.15; P=.02). Moreover, critical eHealth literacy was positively related to consumer health (beta=.30; P<.001) and regular eating habits (beta=.20; P=.002). Thus, hypothesis 1 was partially supported.

The study found that the interactive effect between gender and interactive eHealth literacy was negatively related to balanced diet (beta=−.22; P<.001). The interactive effect between monthly expenses and functional eHealth literacy was positively related to balanced diet (beta=.07; P=.03), although the interactive effect between monthly expenses and critical eHealth literacy was negatively related to balanced diet (beta=−.10; P=.047). Therefore, hypothesis 2 was partially supported.

Discussion

Principal Findings

This study attempted to fully investigate the associations among a broader concept of eHealth literacy and dietary behaviors and to examine the interactive effects of individual factors and the three levels of eHealth literacy on the four aspects of dietary behaviors. We used the IMeHU to explore the associations among individual factors, eHealth literacy, and dietary behaviors. The study found a statistically significant association among the three levels of eHealth literacy and the four aspects

of dietary behaviors. Furthermore, there was an interactive effect between individual factors and eHealth literacy on the balanced diet aspect of dietary behaviors.

This study demonstrated that functional eHealth literacy was negatively related to unhealthy food intake; however, interactive and critical eHealth literacy were not related to unhealthy food intake. This result indicates that the impact of functional eHealth literacy is greater than that of interactive critical eHealth literacy on unhealthy food intake among Taiwanese college students. Previous studies have found that individuals with adequate functional health literacy are less likely to consume fried chicken [31] and sugar-sweetened beverages [11]. Moreover, studies have indicated that targeted intervention strategies that address health literacy, such as quantitative health information guides, are advantageous to reduce sugar-sweetened beverage consumption [32]. Functional eHealth literacy involves basic skills in reading and writing about Web-based health information [17,18]; thus, enabling Taiwanese college students to understand the risks of unhealthy food intake and engage in fewer unhealthy food consumption practices is important.

Consistent with previous studies [19,20], this study showed that interactive eHealth literacy was positively related to the balanced diet and consumer health aspects of dietary behaviors, and critical eHealth literacy was positively related to regular eating habits and consumer health. To our knowledge, this study was the first to investigate the association among the three levels of eHealth literacy and consumer health behaviors. The Taiwanese government has enacted a food education curriculum to improve knowledge of and skills in healthy consumption and to empower consumers to make healthy choices about food and diet. Interactive eHealth literacy involves more advanced cognitive and literacy skills that can be used to actively participate in everyday activities [15] and to promote healthy consumption patterns [33,34]; thus, enriching interactive eHealth literacy might help Taiwanese college students to engage in balanced diet and consumer health practices. Critical eHealth literacy involves the cognitive skills necessary to critically evaluate Web-based health information [16-18] and use such information to make informed decisions about health [16,17]. Critical eHealth literacy allows individuals to evaluate health issues and recognize risks and benefits as well as to advocate for themselves [35], thereby facilitating Taiwanese college students' regular healthy eating habits and consumer health practices.

Perhaps surprisingly, the study found that females with high interactive eHealth literacy were more likely to have a poor balanced diet. Previous studies in Taiwan have also shown female college students' insufficient vegetable intake [28] and a poor balanced diet [7]. It is inferred that social culture plays a central role in the lives of adolescent girls and young women and may influence their female body image and perception of beauty [28]. Although obesity prevalence is quite low among Taiwanese girls, the pressures to be thin still seem to be profound [36]. Gender differences in food choices, therefore, appear to be partly attributable to women's greater weight control involvement [37]. In addition, interactive health communication applications have great potential to improve health, but they may also cause harm [38]. The internet hosts a wide variety of weight loss diets for which individuals may search, but the effects such diets claim may be unconfirmed or exaggerated, false, or even harmful. In addition, more than 15% of internet users have reported feeling overwhelmed and confused by the amount of information available Web-based [39]. Thus, females with high interactive eHealth literacy might be misled by false information and choose unhealthy meals or portions to lose weight, leading to their poor balanced diet. Therefore, better-designed studies are needed to confirm the dieting attitudes and behaviors of females with high interactive eHealth literacy, how eHealth literacy and their body image shape their perceptions, and how body satisfaction may mediate the association between dieting behaviors and health attitudes through their use of Web-based health information.

Consistent with previous studies [23,24], this study found that Taiwanese college students with high monthly expenses were more likely to demonstrate irregular eating habits and consumption of unhealthy food. When living in a student residence, college students become more self-dependent, which also implies that price and budget become increasingly important [40]. College students with a limited budget must choose their food cautiously [24], which can enable regular eating habits and less consumption of snacks and sugary drinks. In contrast, college students with a higher budget might be able to engage in more hedonic eating, resulting in a poor dietary intake that might have a harmful impact on their health and well-being. Paradoxically, this study found that students with higher monthly expenses and higher functional eHealth literacy were more likely to consume a balanced diet, whereas students with higher monthly expenses and higher critical eHealth literacy were less likely to maintain a balanced diet. Therefore, multilevel nutritional interventions may be beneficial to promote healthy eating behavior and dietary intake, particularly among students with high monthly expenses. Schools should provide more nutrition information about food to enhance students' functional eHealth literacy. Future qualitative study is also needed to further examine a broad range of factors that might influence how the level of nutrition-related knowledge and perceptions impact the development of eating patterns among students with high monthly expenses.

Limitations

This study is not without limitations. First, we did not remove responses with little missing or incomplete data because those with very little missing data can still be part of the analysis. Second, unhealthy food intake was measured based on frequency (from seldom to always) rather than an average of kilocalories and consumption per day. Thus, our measure of unhealthy food intake may not accurately indicate the effect of interactive and critical eHealth literacy on unhealthy food intake. Third, the study sample was educated and age restricted in Taiwan. Thus, the findings should not be overgeneralized and should be interpreted in light of the sample's homogeneity. Finally, because the study presented some interesting and paradoxical findings, along with the associations among these factors, eHealth literacy, and dietary behaviors among college students, further studies should examine a broad range of intrapersonal (eg, taste preference, family eating habits, and level of nutrition-related knowledge), interpersonal (eg, peer influence and campus lifestyle), cultural (eg, media and academic activity),

and environmental factors (food availability and food cost) as well as social norms and beliefs that might be influential in determining college students' eating behavior and dietary intake.

Conclusions

This study, to our knowledge, was the first to establish an association among individual factors, the 3 levels of eHealth literacy, and the four aspects of dietary behaviors among college students. This study found that Taiwanese college students with interactive and critical eHealth literacy were more likely to engage in positive dietary behaviors than those with functional eHealth literacy. For the group of students with high monthly expenses, the role of functional eHealth literacy was greater than that of critical eHealth literacy in balanced diet. Moreover, Taiwanese college students with functional eHealth literacy were less likely to engage in unhealthy food intake. These findings have important implications for health educators, who should provide college students with information to help them understand the risks of high-calorie, high-fat, high-salt, or high-sugar diets and thereby reduce their consumption of unhealthy food. As an interactive effect between gender and interactive eHealth literacy on balanced diet was identified, further research is needed to help female college students critically evaluate the credibility and risks of Web-based health information about losing weight and thereby participate in healthy eating practices.

Acknowledgments

This research was financially supported by the Intelligent Electronic Commerce Research Center from The Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education in Taiwan.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Multiple regression analysis of the four aspects of dietary behaviors.
[DOCX File , 24 KB - jmir_v21i11e13140_app1.docx]

References

1. Worsley A. From nutrients to food literacy. J Home Econ Institute Aust 2015;22(3):13-21 [FREE Full text]
2. Deshpande S, Basil MD, Basil DZ. Factors influencing healthy eating habits among college students: an application of the health belief model. Health Mark Q 2009;26(2):145-164. [doi: 10.1080/07359680802619834] [Medline: 19408181]
3. Scott-Sheldon LA, Carey KB, Carey MP. Health behavior and college students: does Greek affiliation matter? J Behav Med 2008 Feb;31(1):61-70 [FREE Full text] [doi: 10.1007/s10865-007-9136-1] [Medline: 17999173]
4. Al-Rethaiaa AS, Fahmy AA, Al-Shwaiyat NM. Obesity and eating habits among college students in Saudi Arabia: a cross sectional study. Nutr J 2010 Sep 19;9:39 [FREE Full text] [doi: 10.1186/1475-2891-9-39] [Medline: 20849655]
5. Racette SB, Deusinger SS, Strube MJ, Highstein GR, Deusinger RH. Weight changes, exercise, and dietary patterns during freshman and sophomore years of college. J Am Coll Health 2005;53(6):245-251. [doi: 10.3200/JACH.53.6.245-251] [Medline: 15900988]
6. Green EJ, Knechtges PL. Food safety knowledge and practices of young adults. J Environ Health 2015 Jun;77(10):18-24. [Medline: 26058218]
7. Luo YF, Yang SC, Chiang CH, Lu CM. Development and validation of a food literacy self-report inventory and investigation of the relationships between food literacy and dietary behavior among college students. Taiwan J Public Health 2018;37(4):407-419. [doi: 10.6288/TJPH.201808_37(4).107029]
8. Parmenter K, Waller J, Wardle J. Demographic variation in nutrition knowledge in England. Health Educ Res 2000 Apr;15(2):163-174 [FREE Full text] [doi: 10.1093/her/15.2.163] [Medline: 10751375]
9. Nutbeam D. Health literacy as a public health goal: a challenge for contemporary health education and communication strategies into the 21st century. Health Promot Int 2000;15(3):259-267. [doi: 10.1093/heapro/15.3.259]
10. Cha E, Kim KH, Lerner HM, Dawkins CR, Bello MK, Umpierrez G, et al. Health literacy, self-efficacy, food label use, and diet in young adults. Am J Health Behav 2014 May;38(3):331-339 [FREE Full text] [doi: 10.5993/AJHB.38.3.2] [Medline: 24636029]
11. Zoellner J, You W, Connell C, Smith-Ray RL, Allen K, Tucker KL, et al. Health literacy is associated with healthy eating index scores and sugar-sweetened beverage intake: findings from the rural Lower Mississippi Delta. J Am Diet Assoc 2011 Jul;111(7):1012-1020 [FREE Full text] [doi: 10.1016/j.jada.2011.04.010] [Medline: 21703379]
12. Karnoe A, Kayser L. How is eHealth literacy measured and what do the measurements tell us? a systematic review. Knowl Manag E-learn 2015;7(4):576-600. [doi: 10.34105/j.kmel.2015.07.038]
13. Norman CD, Skinner HA. eHealth literacy: essential skills for consumer health in a networked world. J Med Internet Res 2006 Jun 16;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]

14. van der Vaart R, Drossaert C. Development of the digital health literacy instrument: measuring a broad spectrum of health 1.0 and health 2.0 skills. J Med Internet Res 2017 Jan 24;19(1):e27 [FREE Full text] [doi: 10.2196/jmir.6709] [Medline: 28119275]
15. Nutbeam D. The evolving concept of health literacy. Soc Sci Med 2008 Dec;67(12):2072-2078. [doi: 10.1016/j.socscimed.2008.09.050] [Medline: 18952344]
16. Hsu WC, Chiang CH, Yang SC. The effect of individual factors on health behaviors among college students: the mediating effects of eHealth literacy. J Med Internet Res 2014 Dec 12;16(12):e287 [FREE Full text] [doi: 10.2196/jmir.3542] [Medline: 25499086]
17. Chiang CH, Yang SC, Hsu WC. Development and validation of the e-health literacy scale and investigation of the relationships between e-health literacy and healthy behavior among undergraduate students in Taiwan. Formosa J Ment Health 2015;28(3):389-420.
18. Paige SR, Stellefson M, Krieger JL, Anderson-Lewis C, Cheong J, Stopka C. Proposing a transactional model of eHealth literacy: concept analysis. J Med Internet Res 2018 Oct 02;20(10):e10175 [FREE Full text] [doi: 10.2196/10175] [Medline: 30279155]
19. Neter E, Brainin E. Association between health literacy, ehealth literacy, and health outcomes among patients with long-term conditions. Eur Psychol 2019 Jan;24(1):68-81. [doi: 10.1027/1016-9040/a000350]
20. Bodie GD, Dutta MJ. Understanding health literacy for strategic health marketing: eHealth literacy, health disparities, and the digital divide. Health Mark Q 2008;25(1-2):175-203. [doi: 10.1080/07359680802126301] [Medline: 18935884]
21. Baker L, Wagner TH, Singer S, Bundorf MK. Use of the internet and e-mail for health care information: results from a national survey. J Am Med Assoc 2003 May 14;289(18):2400-2406. [doi: 10.1001/jama.289.18.2400] [Medline: 12746364]
22. Mitsutake S, Shibata A, Ishii K, Oka K. Associations of eHealth literacy with health behavior among adult internet users. J Med Internet Res 2016 Jul 18;18(7):e192 [FREE Full text] [doi: 10.2196/jmir.5413] [Medline: 27432783]
23. Fu AH, Jien JJ. The impact of college students? attributes on nutrition knowledge, dietary attitudes, and dietary practices in Taiwan. Nutr Sci J 2009;34(4):142-154 [FREE Full text]
24. Chen PY, Lai YM, Chen RY, Sun JL. The relationships between nutritional knowledge, dietary behavior and health status among university students. Journal of Health and Architecture 2016;3(3):52-60.
25. Lang T, Caraher M. Is there a culinary skills transition? Data and debate from the UK about changes in cooking culture. J HEIA 2001;8(2):2-14 [FREE Full text]
26. Davis JN, Ventura EE, Cook LT, Gyllenhammer LE, Gatto NM. LA Sprouts: a gardening, nutrition, and cooking intervention for Latino youth improves diet and reduces obesity. J Am Diet Assoc 2011 Aug;111(8):1224-1230. [doi: 10.1016/j.jada.2011.05.009] [Medline: 21802571]
27. Guttersrud Ø, Petterson KS. Young adolescents' engagement in dietary behaviour - the impact of gender, socio-economic status, self-efficacy and scientific literacy. Methodological aspects of constructing measures in nutrition literacy research using the Rasch model. Public Health Nutr 2015 Oct;18(14):2565-2574. [doi: 10.1017/S1368980014003152] [Medline: 25634262]
28. Fu AH, Jien JJ. Correlation among college students' body shape, dietary habits, and food intake. Nutri Sci J 2008;33(2):55-65 [FREE Full text]
29. Taber KS. The use of Cronbach's alpha when developing and reporting research instruments in science education. Res Sci Educ 2018;48(6):1273-1296. [doi: 10.1007/s11165-016-9602-2]
30. Dormann CF, Elith J, Bacher S, Buchmann C, Carl G, Carré G, et al. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography 2012;36(1):27-46. [doi: 10.1111/j.1600-0587.2012.07348.x]
31. Speirs KE, Messina LA, Munger AL, Grutzmacher SK. Health literacy and nutrition behaviors among low-income adults. J Health Care Poor Underserved 2012 Aug;23(3):1082-1091. [doi: 10.1353/hpu.2012.0113] [Medline: 24212161]
32. Zoellner J, Chen Y, Davy B, You W, Hedrick V, Corsi T, et al. Talking health, a pragmatic randomized-controlled health literacy trial targeting sugar-sweetened beverage consumption among adults: rationale, design & methods. Contemp Clin Trials 2014 Jan;37(1):43-57 [FREE Full text] [doi: 10.1016/j.cct.2013.11.003] [Medline: 24246819]
33. Krause C, Sommerhalder K, Beer-Borst S. Nutrition-specific health literacy: development and testing of a multi-dimensional questionnaire. Ernahrungs Umschau 2016;63(11):214-220 [FREE Full text]
34. Velardo S. The nuances of health literacy, nutrition literacy, and food literacy. J Nutr Educ Behav 2015;47(4):385-9.e1. [doi: 10.1016/j.jneb.2015.04.328] [Medline: 26026651]
35. Mitchell B, Begoray D. Electronic personal health records that promote self-management in chronic illness. Online J Issues Nurs 2010;15(3):1B-10B.
36. Chen L, Fox KR, Haase AM. Body shape dissatisfaction and obesity among Taiwanese adolescents. Asia Pac J Clin Nutr 2008;17(3):457-460 [FREE Full text] [Medline: 18818167]
37. Wardle J, Haase AM, Steptoe A, Nillapun M, Jonwutiwes K, Bellisle F. Gender differences in food choice: the contribution of health beliefs and dieting. Ann Behav Med 2004 Apr;27(2):107-116. [doi: 10.1207/s15324796abm2702_5] [Medline: 15053018]

38. Robinson TN, Patrick K, Eng TR, Gustafson D. An evidence-based approach to interactive health communication: a challenge to medicine in the information age. Science Panel on Interactive Communication and Health. J Am Med Assoc 1998 Oct 14;280(14):1264-1269. [doi: 10.1001/jama.280.14.1264] [Medline: 9786378]
39. Fox S. Online Health Search 2006. 2006. URL: http://goo.gl/Sq5ML4 [accessed 2018-02-18]
40. Deliens T, Clarys P, de Bourdeaudhuij I, Deforche B. Determinants of eating behaviour in university students: a qualitative study using focus group discussions. BMC Public Health 2014 Jan 18;14:53 [FREE Full text] [doi: 10.1186/1471-2458-14-53] [Medline: 24438555]

Abbreviations

DBS: Dietary Behaviors Scale
eHealth: electronic health
eHLS: eHealth Literacy Scale
IMeHU: integrative model of eHealth use
VIF: variance inflation factor

Edited by G Eysenbach; submitted 14.12.18; peer-reviewed by JR Bautista, WC Hsu; comments to author 27.04.19; revised version received 10.06.19; accepted 27.09.19; published 26.11.19.

Please cite as:
Yang SC, Luo YF, Chiang CH
Electronic Health Literacy and Dietary Behaviors in Taiwanese College Students: Cross-Sectional Study
J Med Internet Res 2019;21(11):e13140
URL: http://www.jmir.org/2019/11/e13140/
doi:10.2196/13140
PMID:31769760

Original Paper

Adapting the eHealth Literacy Scale for Carers of People With Chronic Diseases (eHeals-Carer) in a Sample of Greek and Cypriot Carers of People With Dementia: Reliability and Validation Study

Areti Efthymiou[1*], MSc; Nicos Middleton[1*], PhD; Andreas Charalambous[1,2], PhD; Evridiki Papastavrou[1*], PhD

[1]Department of Nursing, Faculty of Health Sciences, Cyprus University of Technology, Limassol, Cyprus

[2]Department of Nursing, Faculty of Health Sciences, University of Turku, Turku, Finland

[*]these authors contributed equally

Corresponding Author:
Areti Efthymiou, MSc
Department of Nursing
Faculty of Health Sciences
Cyprus University of Technology
ZT3, 3rd Floor
15 Vragadinou Street
Limassol, 3041
Cyprus
Phone: 357 25002285
Email: arefthymiou@yahoo.com

Abstract

Background: As the population ages, many more people will be in need of long-term care. According to a recent report by Alzheimer's Disease International and the Karolinska Institute, 84% of people with dementia are cared for at home and 16% in nursing homes. Several Web-based interventions have been developed to assist the work of carers at home. Measuring the levels of electronic health (eHealth) literacy is of top priority to facilitate inclusion of this population and develop training programs to enhance eHealth literacy skills.

Objective: This study aimed to adapt the eHealth Literacy Scale (eHeals) for carers of people with dementia, who speak Greek as their native language and live in Greece and Cyprus, and to test the reliability and validity of the scale for carers.

Methods: The content validity of the eHealth Literacy Scale for Carers of People With Chronic Diseases (eHeals-Carer) was assessed with an expert panel (N=10). A descriptive study with face-to-face interviews among 101 primary carers of people with dementia was conducted. In addition to the eHeals-Carer to assess their perceived eHealth literacy, participants responded to a brief questionnaire regarding characteristics of internet use and provided sociodemographic data. The internal consistency of the tool and the construct validity via an exploratory factor analysis (EFA) were explored.

Results: The Mean Item-Level Content Validity Index (CVI) and Scale-Level CVI Average was 0.93. The participants were mostly women (75.2%, 76/101), aged less than 60 years (67.3%, 68/101) with secondary education. The internal consistency was estimated at a Cronbach alpha of .83. Two factors were extracted from the EFA: information seeking questions 1 to 5 (factor 1) and evaluation questions 6 to 8 (factor 2).

Conclusions: eHeals-Carer is the first perceived eHealth literacy tool adapted for carers of people with dementia. The use of Web-based services available for carers could help them and improve the health care system in the long term. In Greece and Cyprus, there is a lack of services, and improving the digital skills of carers could provide them with the means to support themselves at home and improve care provision.

International Registered Report Identifier (IRRID): RR2-10.2196/resprot.8080

(J Med Internet Res 2019;21(11):e12504) doi:10.2196/12504

KEYWORDS

eHealth; literacy; scales; carers; technology; chronic disease

Introduction

Background

As the population ages, old age diseases are on the rise, that is, many more people will be in need of long-term care in the years to come. In many countries, family and friends usually undertake the role of the carer filling the gap from the lack of organized health and social services, a phenomenon that is more common in Mediterranean and Eastern European regions [1].

According to a recent report by Alzheimer's Disease International and the Karolinska Institute, 84% of people with dementia are cared for at home and 16% in nursing homes [2]. Most carers of people with chronic diseases are aged older than 55 years, and women provide 71% of the annual informal care hours [3,4]. The global number of informal care hours is estimated to be around 6 hours per day or, on an annual basis, 82 billion hours of care. Carers experience stress, making them more vulnerable to infections and memory disorders, and they report a higher use of antidepressants and have high mortality rates [5-7]. The care of people with dementia can be rather demanding, as most patients may develop behavioral disorders in the course of the disease [8]. Carers search for information of the disease prognosis and treatment, services, and support as a way to manage the negative aspects of caregiving and use their social network, friends, families, health providers, and media (newspapers, television, and internet) to do so [9,10].

Carers' Pattern of Use of Web-Based Interventions and the Role of Electronic Health Literacy

Several Web-based interventions have been developed to assist the work of carers at home. They are easy to use and provide quick access to disease-specific information, as in the case of health care websites, psychoeducational platforms, applications, and telehealth and telemonitoring devices [11-13]. In most cases, these services have been provided only during the period of the research intervention, and no further information is provided on their use by carers [14]. According to Chiu and Eysenbach [15], a pattern of use of Web-based interventions made by carers is influenced by several factors such as accessibility, perceived effort, carers' needs (personal skills, social support, carers' beliefs, and years of care), and the style of use. In a modern framework developed to explain factors influencing the design of new technologies based on electronic health (eHealth) literacy level of the users, there is a discussion based on the individual characteristics (being a patient or a carer), the task dimension, and the experience using the technology [16]. Skills in searching, finding, appraising, and applying health information online have also been defined by Norman and Skinner [17], discussing eHealth literacy, which includes the following 6 literacies: traditional, information, media, health, scientific, and computer literacy. The latter 3 (ie, health, scientific, and computer literacy) are categorized according to the authors as context specific. This model has been modified and extended by other researchers [18-20], and a recent definition of eHealth literacy is provided by Bautista [21] and Paige et al [22]. eHealth literacy is redefined and "...involves the interplay of individual and social factors in the use of digital technologies to search, acquire, comprehend, appraise, communicate and apply health information in all contexts of healthcare with the goal of maintaining or improving the quality of life throughout the lifespan." Taking the above into consideration, the individual characteristic, being a carer or a patient, may influence the person's perceived eHealth literacy level. Low health literacy among carers of adults is associated with poorer health provision, care recipient health outcomes, and increased burden [23].

Adapting the eHealth Literacy Scale for Carers of People With Chronic Diseases

There is a lack of published data on eHealth literacy level among carers of people with dementia and adapted or newly developed tools for this purpose.

Norman and Skinner [24] developed the eHealth Literacy Scale (eHeals) to measure the perceived skills that influence the eHealth literacy and consists of 8 items. It was originally tested among 664 adolescents, aged 13 to 21 years, in Canada and showed good metric properties. The scale is easy to administer. The items are short and incorporate a combination of the literacies presented in the Lily model, take no more than 10 min, and assess the way a person searches, assesses, and applies health information online. Even if there is a discussion concerning the lack of Web 2.0 questions [25], at present, it has been translated and used in many different languages and population groups. In the past 5 years, research studies seem to focus on the dimensionality and construct validity of the scale (eg, the number of factors the tool taps on) as well as other related variables such as internet access and use, computer skills, and determinants of eHealth literacy such as age, monthly income, health status, education, and chronic diseases [26-32].

The need for the eHeals to be adapted for the carers population as the eHealth Literacy Scale for carer of chronic diseases (eHeals-Carer), is associated with their caring needs. They usually search information for another person instead of for themselves and their personal health issues, and they are more receptive to technologies that assist them in their caregiving [33,34]. Adapting eHeals items to fit carers' online style of use would facilitate their understanding of the topic and make the questions more comprehensible for their specific needs. This also facilitates their inclusion in the new technological era, as new online tailored services are increasingly provided to carers.

Electronic Health Literacy Among Carers and Available Research in Greece and Cyprus

At the moment, we may only find information on the style of health-related internet use and possible predictors of this type of use made by carers [35,36].

In Greece, recently, a study identified older age and lower education among the main predictors of lower functional eHealth literacy in a Greek-speaking population [32]. We know that in Greece and Cyprus, the main reason for internet nonuse among

older adults is the lack of skills [37,38]. In Greeks and Cypriots, among people aged 65 to 74 years, there is a decrease in internet use from 17.6% in 2012 to 11.1% in 2014 and from 12.7% in 2012 to 6.4% in 2014 for the age group of 75 to 99 years. On the basis of data from the *Internet in Cyprus* report, only 9.6% of the Greek Cypriots search the internet for health information on a weekly basis, and 43% of the sample has never searched the internet for health topics [38].

Objectives

The aim of the study was 2-fold: (1) to identify available validated eHeals as part of a scoping review and (2) to evaluate the validity and reliability of the proposed eHeals for carers among a sample of Greek-speaking carers of people with dementia in Greece and Cyprus.

Methods

Literature Review on Available eHealth Literacy Scale Validations

As part of the validation process, we have searched following the methodology of a scoping review as described in the studies by Arksey and O'Malley and Peters et al [39,40] for relevant validations of eHeals to identify all possible alternatives regarding the different languages, population, statistics, and ratings and any available carers adapted version.

The main research questions of the review are as follows: (1) What type of statistical analysis is used to extract factors for eHeals? (2) How the Web 2.0 problem is handled in existing validations of eHeals? (3) Is there any difference in rating the scale? and (4) Is any eHeals validation for carers available?

We searched for all validations of eHeals in relevant databases (PubMed, CINAHL, MEDLINE, PsycINFO, and Scopus) and gray literature (eScholarship) until December 2018. Keywords used were eHeals and eHealth Literacy Scale.

The studies assessed are based on the following inclusion criteria: (1) the study should be related to the topic of eHealth literacy; (2) the study should be related to the scale reliability and validation; and (3) the study should be published in English

We did not include studies that used eHeals as a measure of eHealth literacy, but no information on validation was provided. The flowchart and related table of results are included in this paper as Multimedia Appendices 1 and 2.

Validation Process of eHeals Carers in Greece and Cyprus

Following the literature review, we designed the validation and adaptation of the eHeals among Greek and Cypriot carers of people with dementia. Permission to use and adaptation of the scale were obtained by the authors [24]. The study followed the validation process as described by the World Health Organization following a double forward and backward translation strategy [41].

As part of the first step, we proceeded with the double forward and backward translation between the original English and Greek. Initially, 2 independent translators, both native speakers of Greek and fluent in English translated the scale into Greek. After comparing and merging the 2 translations into a single Greek translation by consensus, 2 independent back translations into English were derived by an additional set of 2 bilingual translators, 1 care professional and 1 researcher (ie, nurse trainer). In case of disagreement, we employed consensus meeting among the research team members based on expert opinion and existing literature.

In the second step, face validity by the research team followed. During this phase, researchers assessed the available Greek translation of eHeals and if the translated items corresponded to the English version of eHeals. The research team selected the final version in the Greek language and adapted it accordingly by adding a reference to the caregiving concept in every item of the scale. All items were modified accordingly to refer to the health and caregiving issues of a friend/relative, as, for example, in item 1: "I know what health resources are available" adapted to item 1: "I know what resources/information are available on the Internet concerning the health and caregiving issues of my friend/relative." The caregiving issues on the scale are explained as the practical, financial, legal issues and information about the disease and available services. In the case of items 2, 3, and 4, we also added short clarification to facilitate understanding. Modifications of the scale are available in Table 1.

The content validity of the adapted items in the Greek language was assessed by a panel of experts in the field of eHealth and dementia or older people. Following this process, the questionnaire was piloted in 25 carers. Finally, the internal consistency of the final version of the Greek-adapted scale was tested among a sample of primary carers, and construct validity was followed with exploratory factor analysis (EFA).

Table 1. eHeals-Carer (eHealth Literacy Scale for Carers of People With Chronic Diseases) items: item difficulty, corrected item-total correlation, and factor loading.

Questions per factor	Mean (SD)	Median	Corrected item-total correlation	Factor loadings
Factor 1				
Item 1: "I know what resources/information are available on the Internet concerning the health and caregiving issues of my friend/relative (practical, financial, legal issues, information about the disease and available services)."	3.51 (0.93)	4	0.48	0.485
Item 2: "I know where to find helpful information on the Internet concerning the health and caregiving of my friend/relative (e.g. which websites I will search)."	3.35 (1.06)	4	0.59	0.540
Item 3: "I know how to find helpful information on the Internet concerning the health and caregiving of my friend/relative (e.g concerning the process: google search)."	4.08 (0.82)	4	0.55	0.735
Item 4: "I know how to use the Internet to answer my questions about the health and caregiving of my friend/relative (e.g how to ask in order to receive a proper reply to my question)."	3.83 (1)	4	0.53	0.656
Item 5: "I know how to use the information about the health and caregiving of my friend/relative I find on the Internet to help me (practical, financial, legal issues, information about the disease and available services)."	3.75 (0.85)	4	0.55	0.500
Total	18.49 (19)	19	—[a]	—
Factor 2				
Item 6: "I have the skills I need to evaluate the resources/information I find on the Internet concerning the health and caregiving of my friend/relative."	3.70 (1.05)	4	0.59	0.756
Item 7: "I can tell high quality resources/information from low quality resources/information on the Internet concerning the health and caregiving of my friend/relative."	3.75 (1)	4	0.59	0.731
Item 8: "I feel confident in using information from the Internet to make decisions concerning the health and caregiving of my friend/relative."	3.30 (1.08)	3	0.57	0.595
Total	10.77 (2.62)	11	—	—
Total scores from both factors	29.27 (5.30)	29	—	—

[a]Not applicable.

Recruitment

Recruitment Panel of Experts for the Content Validity Index

To proceed with the content validity index, we invited 10 experts to reply to the content validity of the questionnaire. The experts were invited because of their work on eHealth and/or dementia domain. Of 10 experts, 8 were health professionals: 3 health care professionals, nurses, and psychologists working in the field of technology (robotics and digital literacy of older people), 1 member of the Greek team of the European Health Literacy Survey, and 4 health care professionals working in dementia care. The remaining 2 were information technology experts working in the field of eHealth.

Recruitment of Primary Carers

The data collection of primary carers was made in the framework of the research protocol for "the Association of Health Literacy and Electronic Heath Literacy with Self-Efficacy, Coping and Caregiving Perceptions Among Carers of People with Dementia: Research Protocol for a Descriptive Correlational Study" [42].

The final sample of the protocol was estimated with 95% power and a type 1 error of 5% to 168 primary carers. All questionnaires were pilot tested in 25 primary carers of people with dementia [43].

The validation of eHeals adapted for carers proceeded with a convenience sample of 101 carers from Greece and Cyprus, based on the subject-to-item ratio 10:1 [43-45]. Participation in the study was voluntary, and the recruitment of the sample lasted for 1 year. Eligibility criteria were broad and included being a carer of a person with dementia, speaking Greek, and being aged older than 18 years. Researchers approached carers at Dementia Day Care Centers in Athens, Greece, and Limassol, Cyprus, or during training courses and public awareness campaign events directed to carers of people with dementia. In the case of Dementia Centers, the scientific supervisors assisted the researcher to arrange the appointment at the time of the day

that carers were available. In the case of public events, the researcher distributed leaflets, and carers expressed their interest in participating. The researcher arranged a face-to-face survey appointment to administer the questionnaire.

Measures

The measures were as follows:

- Content Validity Index [46]: all expert panel participants received the questionnaire adapted for carers in the Greek language and assessed item phrasing, simplicity by commenting on every item and relevance on a 4-point scale: not relative, somehow relative, quite relative, and relative.
- Carers replied to the Greek version of eHeals-Carer, which includes 8 items, each with a 5-point response scale from 1 (strongly disagree) to 5 (strongly agree). As shown in Table 1, all 8 items were adapted accordingly to specifically refer to the caregiving role.
- Carers also provided the following basic sociodemographic information: gender, age, education level (based on the international standard classification of education), employment status, carers' relationship, living status, and being supported by a secondary carer or not), and replied to a series of questions with regard to internet use, either personal or dementia-specific online use. As part of the sociodemographic information, we have used a visual analog scale for measuring the socioeconomic position, Ladder questionnaire [47,48]. The participants were asked to assess where they stood on a ladder in comparison with other people in Greece or Cyprus, given that in the bottom of the scale are the people with the worst profession or unemployment, least money, and lowest education.

Data Analysis

In content validity, we reported the following 3 indexes: (1) Mean Item-Level Content Validity Index (Mean I-CVI), measuring the proportion of relative and very relative responses of the items; (2) Scale-Level Content Validity Index Average (S-CVI/Ave), measuring the average score of the responses of quite relevant and very relevant of every expert; and (3) the Scale Content Validity Index Universal Agreement (S-CVI/UA), measuring all items that all raters assessed as quite or highly relative. As scale CVI, we usually consider the S-CVI/Ave because the S-CVI/UA decreases as the number of raters increases [46].

The internal consistency of the scale was assessed with a Cronbach alpha, and the dimensionality of the scale was explored with EFA. This was the first time that the scale was validated in Greek among carers, and dimensions were not hypothesized before the validation. Confirmatory factor analysis (CFA) will be calculated with the total sample of the study protocol based on the EFA findings.

Ethical Consideration

The Cyprus National Ethical Committee (ΕΕΒΚ ΕΠ 2016.01.151) and the Cyprus Commissioner for Personal Data Protection (3.28.460) approved the study. As the study was conducted in 2 countries, the study protocol also received approval by the Scientific Committee of Alzheimer Athens Association (March 17, 2017).

Results

Results of Literature Review on Available eHealth Literacy Scale Validations

According to the first step of the validation process, we conducted a review to identify all possible eHeals validations to decide on the methodology and avoid any replication of existing measures for this specific population.

The scale has been validated and adapted in many different languages and population groups, using either convenient sample recruitment strategies or randomized recruitment techniques (as random telephone dialing). In the last 3 years, the validation studies of the specific tool were increased, showing a tendency toward eHealth literacy research. Only in 1 study from Slovenia did we find the validation of an extended version of 20 items (6 factors) including the Web 2.0 parameter as discussed earlier by Norman [49,50]. In 21 cases, the authors preferred a combination of the original scale adding questions to assess health-related internet use and internet use in general [17,25,27-29,31,51-64]. The reliability in the majority of the studies was quite high, that is, over 0.80. The lowest reliability was presented in a student sample in Bangladesh and in the 6 dimensions of the Slovenian version [50,61]. In 6 of 26 studies, the sample recruitment focused on older adults [25,52,59,65-67].

A series of studies have identified or confirmed the unidimensionality of the eHeals [25,30,31,57,68-70]. However, the latest studies seem to propose either a 2-factor model or a 3-factor model [27-29,52,54,59,62,67]. The study by Soellner et al [64] was one of the first to propose a 2-factor model with an information seeking (questions 1-5 and 8) and an information appraisal (questions 6 and 7) component. This model was later confirmed by Diviani et al [28]. Subsequent studies also supported a 2-factor model, yet with a different set of questions, for example, the first 4 questions tapping on factor 1 and the last 4 questions on factor 2 [27,29]. With regard to the 3-factor model, the most commonly accepted dimensions are as follows: awareness (questions 1 and 2), skills (questions 3-5), and evaluation (questions 6-8). Paige et al [63] proposed a 3-factor model with a different categorization, which, instead of skills and evaluation, includes information seeking (questions 3 and 4) and information engagement (questions 5-8).

In almost all cases, the scoring system distinguished between high and low scores without providing information for a medium level. In 12 papers, the level was calculated by summarizing all items, and in 4 validation studies, the level was calculated by summing up all items and dividing the score with the number of the scale or of the factor. The highest score of eHeals among the studies included in this review is presented in the study by Chung and Nahm [65] for a sample of 886 adults, with a mean age of 62 years and eHeals literacy mean score of 30.94 (SD 6).

In 5 studies, the researchers used a principal component analysis (PCA), in 11 cases EFA, in 8 studies CFA, and in 3 studies either PCA or EFA and then CFA (Multimedia Appendix 2).

In 4 studies, they followed item response theory and Rasch modeling.

This review provided the basis for our validation study. On the basis of the above results, the discussion for the use of classical test theory and item response theory in behavioral and social science [71], and the aim of our study (to adapt an already developed short scale), we decided to follow the classical test theory validation and the use of EFA. As there were many available validations providing different dimensions, we decided to explore the dimensions in this target group and confirm these factors in a larger study sample of carers.

Our decision to adapt for a specific population was in accordance with the measurement modifications for diverse populations [72]. The reasons for modifying this scale were as follows: (1) carers were a different population from the one that participated in the development of the original scale; (2) the scale lacks the caregiving concept that carers would be related to; and (3) if the eHeals was used as it is, there might be a misinterpretation of the items through the caregiving filter. To proceed with the adaptation of the eHeals, we followed an extensive literature review on the eHealth literacy research among carers and older people. Carers' research on eHealth literacy was limited, but we encountered valuable information on the internet use among carers of frail older people and people with dementia. On the basis of this research, we were able to understand how carers may use the internet in relation to caregiving. They mostly searched for disease-specific information, services for the patients, practical issues, and legal and financial issues and to communicate through emails and chat sites [73-75]. In this regard, we decided to proceed with the context-specific modifications of the eHeals as has been discussed in the following subsections.

Content Validity of eHealth Literacy Scale Carers in Greek

Mean I-CVI and S-CVI/Ave was 0.93 in both cases. S-CVI/UA was 0.60.

Experts made no further comment on the phrasing of the scale, apart from 3 comments on 3 different items (items 1, 2, and 9), that did not change the final meaning of these items.

Demographic Information of Primary Carers

As part of the reliability and construct validity, our sample comprised primary carers, mostly women (75.2%, 76/101), caring for their parents (61.3%, 62/101) living in the same household (61.3%, 62/101), aged younger than 60 years (67.3%, 68/101), having completed 12 years of education or more (92.0%, 93/101), mostly unemployed or pensioners (62.3%, 63/101), and receiving assistance from a secondary carer (78.2%, 79/101). Detailed demographics are presented in Table 2. Socioeconomic position was assessed with the use of the ladder figure questionnaire with 10 steps, providing a mean score of 5.8.

Table 2. Demographic information of the carers sample (N=101).

Characteristics	Value, n (%)
Gender	
Women	76 (75)
Men	25 (25)
Age (years)	
<59	68 (67)
60-79	33 (33)
>80	0 (0)
Education	
No primary education (ISCED[a], level 0)	0 (0)
Primary education (ISCED, level 1)	8 (8)
Secondary education (ISCED, levels 2-4)	54 (53)
Tertiary education (ISCED; levels 5.1, 5.2, and 6)	39 (39)
Employment status	
Employed	38 (38)
Unemployed (including pensioners)	63 (62)
Carers' relationship	
Caring for parent	62 (61)
Caring for spouse	28 (28)
Caring for other (relative/friend/neighbor)	11 (11)
Secondary carer support	
Yes	79 (78)
No	22 (22)
Living status	
Together with person with dementia	62 (61)
Living in other's house	39 (39)
Most frequent internet use for carers	
Search of information	40 (43)
Reading news	15 (16)
Entertainment (movies and music)	12 (13)
Social networks	8 (9)
Emails	9 (10)
Professional reasons	8 (9)

[a]ISCED: International Standard Classification of Education.

Internet Use Characteristics

Of 101 participants, 92 used the internet with the more frequent reason of private internet use: *searching for information on different topics*. Of all participants, 97.0% (98/101) visited websites; 76.2% (77/101) used social networks, such as Facebook, Twitter, and LinkedIn; 81.1% (82/101) used email to communicate; 83.1% (84/101) used interactive services (eg, Viber, Skype, forums, and chatrooms); and only 42.5% (43/101) accessed electronic learning (eLearning) courses.

In the questions regarding online search of dementia-specific information such as disease information, practical issues, legal information, and available services, almost all participants 90.0% (91/101) stated that they had accessed online dementia resources and mostly websites. Almost half of the participants (40.5%, 41/101) had used social networks, and 42% (42/101) had used email to communicate and searched for information with other carers, family, and health professionals. The use of interactive services and eLearning courses were the least preferred resources to communicate and receive information or

training with 32.6% (33/101) and 12% (11.8/101) users equivocally.

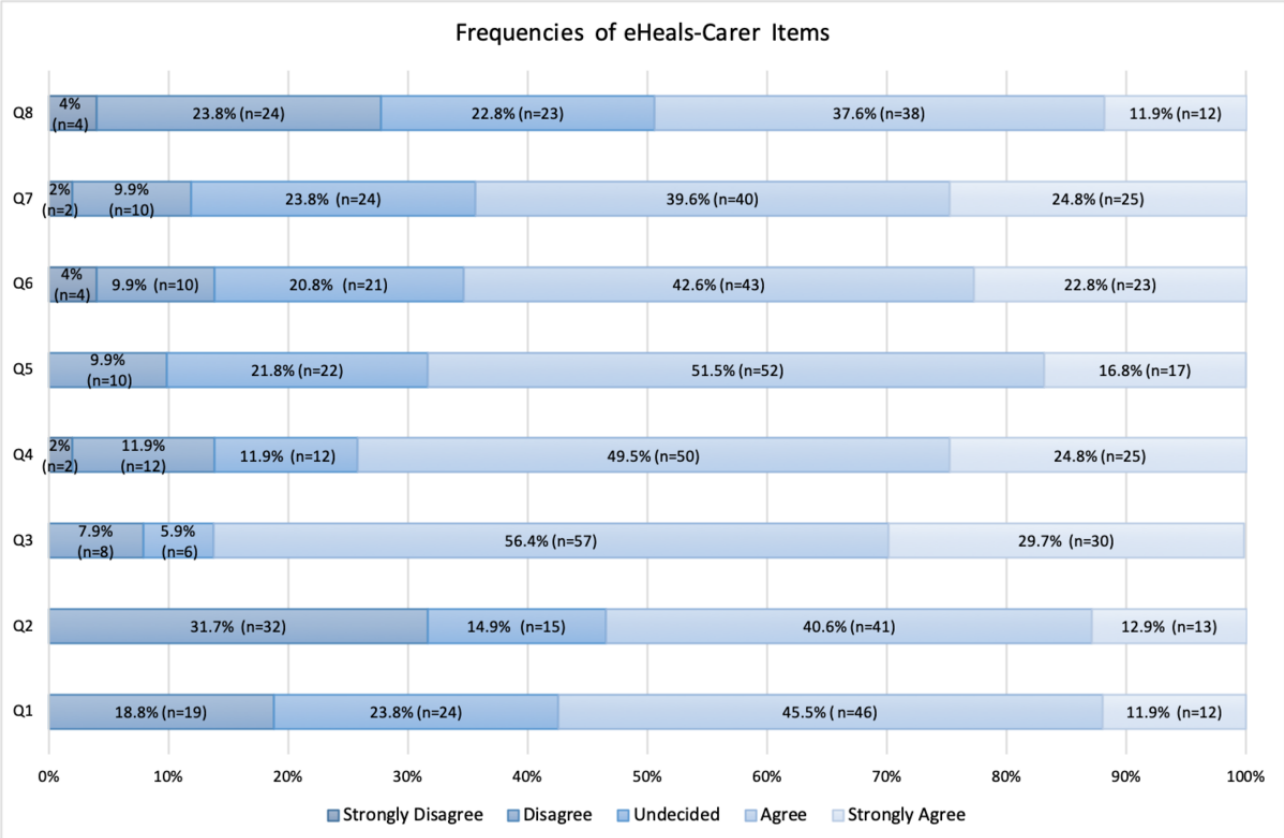
Among all participants, 51.4% (52/101) used a mobile phone to access information for dementia care or to communicate with other carers or health care professionals. Adding to the above result, of 52 participants who have used the internet on their mobile phone, 86% (45/52) have accessed websites, 54% (28/52) accessed social networks, 39% (20/52) used emails, 42% (22/52) used other interactive services, and 5% (3/52) used eLearning services through their mobile phone.

Reliability

Internal consistency of the scale was measured with Cronbach alpha of .83. All items appeared important with item-total correlations ranging between .48 and .59. In all cases, the Cronbach alpha was lower if any of the items was removed.

The items with the highest frequency of replies of agreement (agree and strongly agree) were item 3 "I know how to find helpful information on the Internet concerning health and caregiving of my friend/relative (e.g. concerning the process: google search)," item 4 "I know how to use the Internet to answer my questions about the health and caregiving of my friend/relative (e.g. how to ask in order to receive a proper reply to my question)," and item 5 "I know how to use the information about the health and caregiving of my friend/relative I find on the Internet to help me (practical, financial, legal issues, information about the disease and available services)." Item 8 "I feel confident about using information from the Internet to make decisions concerning the health and caregiving of my friend/relative" had the lowest scores of agreement (Figure 1). This was also confirmed by mean scores of every item of the scale as presented in Table 1. The total mean score of the scale eHeals-Carer was 29.27 (SD 5.30).

Figure 1. Frequencies of responses of eHeals-Carer (eHealth Literacy Scale for Carers of People With Chronic Diseases) items.



Construct Validity

The dimensionality of the scale was explored in EFA, principal axis factoring with Varimax rotation. Kaiser-Meyer-Olkin measure sampling adequacy was 0.80, and the Bartlett test of sphericity was statistically significant (χ^2_{28}=261.5 P<.001). Overall, 2 factors with eigenvalue greater than 1 were extracted, with the first factor explaining 24% of the variance and the second factor 23% (rotation sums of square loadings). After Varimax rotation, a clear structure was revealed with no cross-loadings. Items 1 to 5 loaded on the first factor and seem to tap on the *information seeking* aspect of eHealth literacy. Items 6 to 8 loaded on the second factor and tapped on the *evaluation* aspect of eHealth literacy. Reliability analysis for factor 1 provided a Cronbach alpha of .77 (mean 18.48 [SD 3], median 19), and for factor 2, a Cronbach alpha of .78 (mean 10.77 [SD 2.62], median 11).

Discussion

Principal Findings

We searched the literature to identify all possible validations of the eHeals and to check if there was any adapted version for this population. We adapted and validated the scale for carers, resulting in a scale with high Mean I-CVI (0.93) and high reliability (0.83). The data analysis supported 2 factors:

information seeking and evaluation. The first factor includes the 5 items of eHeals 1 to 5, and the second factor includes 3 items 6 to 8. In the literature, we identify different categories derived from the analysis of eHeals including awareness (1 and 2), skills (3-5), information seeking (1-5 and 8 or 3-4), information appraisal (6 and 7), information engagement (5-8), and evaluation (6-8). We have also identified 2 factors related to seeking and appraisal skills as in the case of Soellner et al [54], but with a different combination of the eHeals items for the 2 dimensions. This difference, from other researchers, might derive from the cultural adaptation of the tool. In item 5 "I know how to use the information about the health and caregiving of my friend/relative I find on the Internet to help me (practical, financial, legal issues, information about the disease and available services)" was perceived as a competence/skill item on how to do rather than as an item for evaluating the information.

In eHeals, as initially developed by Norman and Skinner, more than 1 literacy is included per item of eHeals [17,53]. For example, traditional, information, computer, and health literacy are included in all items of the scale. Media and scientific literacy can be identified in the evaluation subscale [53]. We adapted the short-scale 8-item eHeals for carers to investigate carers' eHealth literacy levels. In this adaptation, we consider the different needs of carers regarding health and eHealth literacy skills. According to a recent scoping review, carers' levels of health literacy are considered adequate, even if they largely depend on the scale used [23]. Carers are the people who manage the communication with the health care providers and the care recipient, manage support services for the dependent person, and make health-related decisions. We also know from previous studies that carers' health literacy levels and eHealth literacy skills may vary according to the person's characteristics: being a carer or not, as this has been identified for the health-related internet use in this population [36]. Carers report higher levels of health literacy in comparison with the care recipients [23]. They usually search for health-related information for the cared-for person and use the internet to find information about the disease prognosis and treatment, legal and financial issues, practical issues, and communication [34,36,73]. Online information and services are important for the health self- management [9]. This is also confirmed in a study by Anderson et al and the analysis of 2345 carers' posts in 9 websites. Researchers have categorized posts in 4 topics: social support—communication and inclusion, search of information, sharing of memories with the person with dementia, and sharing information with other carers [76].

In Greece and Cyprus, carers are the core element in the care provision of people with dementia, covering the lack of tailored services by the National Health System [77,78]. The development of eHealth tools has been promising in this area, assisting carers in everyday tasks, but still much needs to be done to increase the use of these tools by carers. As a first step, we need to investigate the eHealth literacy levels of carers by using a short, easy-to-comprehend tool. In this study, we adapted the eHeals questionnaire to mirror the carers' role as an effort to provide this adapted tool to carers in Greece and Cyprus. In Greece, Xesfingi and Vozikis [32] assessed the eHealth literacy level in a sample of 1064 citizens, ages ranging from 15 years to older than 80 years, with older people and the less educated to be less eHealth literate. In Cyprus, there is no available literature measuring eHealth literacy levels among older people or carers.

We consider that this scale assists in the assessment of eHealth literacy level of carers in 2 ways. Firstly, in practice, as the health care system, not-for-profit organizations, and academic institutions could develop tailored programs for the online needs of the carers. In this way, carers may improve the way they access and evaluate dementia-specific information or information regarding their health. Secondly, in research, as we provide a validated tool for use in future studies investigating the determinants of eHealth literacy, its association with the burden and other aspects of the caregiving role, as well as a process outcome measure in intervention studies targeting eHealth literacy. In this way, eHealth inequalities may be decreased, as carers improve the management of the disease and their burden because of a better use of the available Web-based services.

Finally, through the validation process in this diverse population, we identified culturally specific issues related to the understanding of the items of the first-dimension *seeking information*, and we consider important in future research on the development and validation of eHealth literacy tools that researchers include short exemplars to facilitate understanding of the *how to* items when related with internet users' skills.

Limitations

Carers of people with dementia in this study are considered a convenient sample. Participation rate did not exceed 31% as revealed in the piloting phase of the study protocol. Carers in Greece and Cyprus were not easy to identify if they had not attended a dementia center. As a consequence, the final sample included in this validation was small. The study should be repeated in a larger sample, among carers of patients of other chronic diseases and could be used for cross-country comparisons between Greek and Greek-Cypriot carers.

Even if the eHeals has been adapted for carers, no item about Web 2.0 has been added in the 8-item scale. We only added it in the supplementary section of the internet use characteristics [49]. Carers use the internet to interact with health care professionals and other carers [79-81]. This type of internet use (interaction with social networks: forums and chatrooms) is not depicted in this scale, making this adapted version limited but convenient for use in large study protocols when there is a need of a short tool with high reliability and validity for measuring eHealth literacy among carers.

Conclusions

The validation of eHeals-Carer provides the first questionnaire measuring perceived eHealth literacy skills adapted to carers. At the moment, there is no other scale measuring eHealth literacy levels for carers available. The development of new tools on eHealth literacy measuring functional aspects adapted to specific needs seems to be the next step in this research area. Carers of people with dementia, in the majority, are people aged older than 50 years, children, or spouses, with low use of

care-specific Web-based services. The use of the online services available for carers could facilitate the carers and the long-term health care system. In Greece and Cyprus, there is a lack of services for carers, and by improving their digital skills, we could provide them with the means to support themselves and improve the care they provide. With the increased offer of Web-based services tailored for carers, the improvement of their digital skills will become more demanding in the years to come. Furthermore, public and private services in Greece and Cyprus are updating their service systems to be following technological progress. In this era, carers can be included if we provide them with adequate and appropriate eHealth literacy training programs.

Acknowledgments

Funding for publication was granted by the Cyprus University of Technology Open Access Author Fund. The authors would like to thank all the experts contributed in the process of the validation of the scale and the supervisors and social workers of Dementia Centres of Athens Association of Alzheimer Disease, Greek Association of Alzheimer's Disease in Thessaloniki, IASI Dementia Centre in Glyfada, and Ithaki Alzheimer Association in Cyprus for their availability to use their facilities of the research and facilitate access to their members.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Scoping review flowchart.
[PDF File (Adobe PDF File), 38 KB - jmir_v21i11e12504_app1.pdf]

Multimedia Appendix 2
Scoping review validation results of eHeals (Electronic Health Literacy Scale).
[PDF File (Adobe PDF File), 65 KB - jmir_v21i11e12504_app2.pdf]

References

1. Triantafillou J, Naiditch M, Repkova K, Stiehr K, Carretero S, Emilsson T, et al. Research Gate. Informal Care in the Long-Term Care System URL: https://www.researchgate.net/profile/Judy_Triantafillou/publication/242658253_Informal_care_in_the_long-term_care_system/links/02e7e529488dd69874000000/Informal-care-in-the-long-term-care-system.pdf [accessed 2019-10-07]
2. Wimo A, Gauthier S, Prince M. Alzheimer's Disease International. 2018. Global Estimates of Informal Care URL: https://www.alz.co.uk/adi/pdf/global-estimates-of-informal-care.pdf [accessed 2019-10-07]
3. CORDIS | European Commission. 2006. Services for Supporting Family Carers of Elderly People in Europe : Characteristics, Coverage and Usage URL: https://cordis.europa.eu/project/rcn/67322/factsheet/en [accessed 2019-10-07]
4. Lüdecke D, Bien B, McKee K, Krevers B, Mestheneos E, di Rosa M, et al. For better or worse: factors predicting outcomes of family care of older people over a one-year period. A six-country European study. PLoS One 2018;13(4):e0195294 [FREE Full text] [doi: 10.1371/journal.pone.0195294] [Medline: 29614108]
5. Pinquart M, Sörensen S. Differences between caregivers and noncaregivers in psychological health and physical health: a meta-analysis. Psychol Aging 2003 Jun;18(2):250-267. [doi: 10.1037/0882-7974.18.2.250] [Medline: 12825775]
6. Pinquart M, Sörensen S. Associations of caregiver stressors and uplifts with subjective well-being and depressive mood: a meta-analytic comparison. Aging Ment Health 2004 Sep;8(5):438-449. [doi: 10.1080/13607860410001725036] [Medline: 15511742]
7. Vitaliano PP, Zhang J, Scanlan JM. Is caregiving hazardous to one's physical health? A meta-analysis. Psychol Bull 2003 Nov;129(6):946-972. [doi: 10.1037/0033-2909.129.6.946] [Medline: 14599289]
8. Tible OP, Riese F, Savaskan E, von Gunten A. Best practice in the management of behavioural and psychological symptoms of dementia. Ther Adv Neurol Disord 2017 Aug;10(8):297-309 [FREE Full text] [doi: 10.1177/1756285617712979] [Medline: 28781611]
9. Allen F, Cain R, Meyer C. Seeking relational information sources in the digital age: a study into information source preferences amongst family and friends of those with dementia. Dementia (London) 2018 Jan 1:1471301218786568. [doi: 10.1177/1471301218786568] [Medline: 29999410]
10. Washington KT, Meadows SE, Elliott SG, Koopman RJ. Information needs of informal caregivers of older adults with chronic health conditions. Patient Educ Couns 2011 Apr;83(1):37-44. [doi: 10.1016/j.pec.2010.04.017] [Medline: 20452165]
11. Barbabella F, Poli A, Andréasson F, Salzmann B, Papa R, Hanson E, et al. A web-based psychosocial intervention for family caregivers of older people: results from a mixed-methods study in three European countries. JMIR Res Protoc 2016 Oct 6;5(4):e196 [FREE Full text] [doi: 10.2196/resprot.5847] [Medline: 27713113]

12. Cristancho-Lacroix V, Wrobel J, Cantegreil-Kallen I, Dub T, Rouquette A, Rigaud A. A web-based psychoeducational program for informal caregivers of patients with Alzheimer's disease: a pilot randomized controlled trial. J Med Internet Res 2015 May 12;17(5):e117 [FREE Full text] [doi: 10.2196/jmir.3717] [Medline: 25967983]
13. Papa R, Piccini F, Onorati G, Lamura G, Boccaletti L, Manattini A, et al. Research Gate. 2017. APPs4Carers: selection of applications and mobile websites for informal carers URL: https://www.researchgate.net/publication/318529006_APPs4Carers_selection_of_applications_and_mobile_websites_for_informal_carers [accessed 2019-10-07]
14. Wasilewski MB, Stinson JN, Cameron JI. Web-based health interventions for family caregivers of elderly individuals: a Scoping Review. Int J Med Inform 2017 Jul;103:109-138. [doi: 10.1016/j.ijmedinf.2017.04.009] [Medline: 28550996]
15. Chiu TM, Eysenbach G. Theorizing the health service usage behavior of family caregivers: a qualitative study of an internet-based intervention. Int J Med Inform 2011 Nov;80(11):754-764. [doi: 10.1016/j.ijmedinf.2011.08.010] [Medline: 21958550]
16. Norgaard O, Furstand D, klokker L, Karnoe A, Batterham R, Kayser L, et al. The e-health literacy framework: A conceptual framework for characterizing e-health users and their interaction with e-health systems. Knowl Manage E-learn 2015;7(4):522-540 [FREE Full text]
17. Norman CD, Skinner HA. eHealth literacy: essential skills for consumer health in a networked world. J Med Internet Res 2006 Jun 16;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
18. Chan CV, Kaufman DR. A framework for characterizing eHealth literacy demands and barriers. J Med Internet Res 2011 Nov 17;13(4):e94 [FREE Full text] [doi: 10.2196/jmir.1750] [Medline: 22094891]
19. Gilstad H. Toward a Comprehensive Model of eHealth Literacy. In: Proceedings of the Practical Aspects of Health Informatics 2014. 2014 Presented at: PAHI'14; May 19-20, 2014; Trondheim, Norway URL: http://ceur-ws.org/Vol-1251/paper7.pdf
20. Koopman RJ, Petroski GF, Canfield SM, Stuppy JA, Mehr DR. Development of the PRE-HIT instrument: patient readiness to engage in health information technology. BMC Fam Pract 2014 Jan 28;15:18 [FREE Full text] [doi: 10.1186/1471-2296-15-18] [Medline: 24472182]
21. Bautista JR. From solving a health problem to achieving quality of life: redefining eHealth literacy. J Lit Technol 2015;16(2):33-54 [FREE Full text]
22. Paige SR, Stellefson M, Krieger JL, Anderson-Lewis C, Cheong J, Stopka C. Proposing a transactional model of eHealth literacy: concept analysis. J Med Internet Res 2018 Oct 2;20(10):e10175 [FREE Full text] [doi: 10.2196/10175] [Medline: 30279155]
23. Yuen EY, Knight T, Ricciardelli LA, Burney S. Health literacy of caregivers of adult care recipients: a systematic scoping review. Health Soc Care Community 2018 Mar;26(2):e191-e206. [doi: 10.1111/hsc.12368] [Medline: 27426731]
24. Norman CD, Skinner HA. eHEALS: the eHealth literacy scale. J Med Internet Res 2006 Nov 14;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
25. van der Vaart R, van Deursen AJ, Drossaert CH, Taal E, van Dijk JA, van de Laar MA. Does the eHealth Literacy Scale (eHEALS) measure what it intends to measure? Validation of a Dutch version of the eHEALS in two adult populations. J Med Internet Res 2011 Nov 9;13(4):e86 [FREE Full text] [doi: 10.2196/jmir.1840] [Medline: 22071338]
26. Choi NG, Dinitto DM. The digital divide among low-income homebound older adults: internet use patterns, eHealth literacy, and attitudes toward computer/Internet use. J Med Internet Res 2013 May 2;15(5):e93 [FREE Full text] [doi: 10.2196/jmir.2645] [Medline: 23639979]
27. Dashti S, Peyman N, Tajfard M, Esmaeeli H. E-Health literacy of medical and health sciences university students in Mashhad, Iran in 2016: a pilot study. Electron Physician 2017 Mar;9(3):3966-3973 [FREE Full text] [doi: 10.19082/3966] [Medline: 28461871]
28. Diviani N, Dima AL, Schulz PJ. A psychometric analysis of the Italian version of the eHealth literacy scale using item response and classical test theory methods. J Med Internet Res 2017 Apr 11;19(4):e114 [FREE Full text] [doi: 10.2196/jmir.6749] [Medline: 28400356]
29. Gazibara T, Cakic J, Cakic M, Pekmezovic T, Grgurevic A. eHealth and adolescents in Serbia: psychometric properties of eHeals questionnaire and contributing factors to better online health literacy. Health Promot Int 2019 Aug 1;34(4):770-778. [doi: 10.1093/heapro/day028] [Medline: 29800141]
30. Koo M, Norman C, Chang HM. Psychometric evaluation of a Chinese version of the eHealth literacy scale (eHEALS) in school age children. Int Electron J Health Educ 2012;15:29-36 [FREE Full text]
31. Neter E, Brainin E. eHealth literacy: extending the digital divide to the realm of health information. J Med Internet Res 2012 Jan 27;14(1):e19 [FREE Full text] [doi: 10.2196/jmir.1619] [Medline: 22357448]
32. Xesfingi S, Vozikis A. eHealth literacy: in the quest of the contributing factors. Interact J Med Res 2016 May 25;5(2):e16 [FREE Full text] [doi: 10.2196/ijmr.4749] [Medline: 27226146]
33. Alwan M, Orlov L, Schulz R, Vuckovic N. National Alliance for Caregiving. 2011. E-Connected Family Caregiver: Bringing Caregiving into 21st Century URL: https://www.caregiving.org/data/FINAL_eConnected_Family_Caregiver_Study_Jan%202011.pdf [accessed 2019-10-07]
34. Fox S, Brenner J. Pew Internet. 2012. Family Caregivers Online URL: https://www.pewinternet.org/wp-content/uploads/sites/9/media/Files/Reports/2012/PIP_Family_Caregivers_Online.pdf [accessed 2019-10-07]

35. Blackburn C, Read J, Hughes N. Carers and the digital divide: factors affecting Internet use among carers in the UK. Health Soc Care Community 2005 May;13(3):201-210. [doi: 10.1111/j.1365-2524.2005.00547.x] [Medline: 15819741]
36. Kim H. Understanding internet use among dementia caregivers: results of secondary data analysis using the US caregiver survey data. Interact J Med Res 2015 Feb 23;4(1):e1 [FREE Full text] [doi: 10.2196/ijmr.3127] [Medline: 25707033]
37. Hellenic Statistical Authority. 2011. 2011 Population and Housing Census URL: http://www.statistics.gr/documents/20181/1215267/A1602_SAM05_DT_DC_00_2011_02_F_EN.pdf/62003ed0-2e33-4d0e-9cda-1a262f2b8982 [accessed 2019-10-07]
38. Millioni D, Stylianos S. Research Gate. 2015. The Internet in Cyprus 2014-Final Research Report URL: https://www.researchgate.net/publication/296333618_The_Internet_in_Cyprus_2014-Final_Research_Report [accessed 2019-10-07]
39. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol Theory Pract 2005;8(1):19-32. [doi: 10.1080/1364557032000119616]
40. Peters MD, Godfrey CM, Khalil H, McInerney P, Parker D, Soares CB. Guidance for conducting systematic scoping reviews. Int J Evid Based Healthc 2015 Sep;13(3):141-146. [doi: 10.1097/XEB.0000000000000050] [Medline: 26134548]
41. Tsang S, Royse CF, Terkawi AS. Guidelines for developing, translating, and validating a questionnaire in perioperative and pain medicine. Saudi J Anaesth 2017 May;11(Suppl 1):S80-S89 [FREE Full text] [doi: 10.4103/sja.SJA_203_17] [Medline: 28616007]
42. Efthymiou A, Middleton N, Charalambous A, Papastavrou E. The association of health literacy and electronic health literacy with self-efficacy, coping, and caregiving perceptions among carers of people with dementia: research protocol for a descriptive correlational study. JMIR Res Protoc 2017 Nov 13;6(11):e221 [FREE Full text] [doi: 10.2196/resprot.8080] [Medline: 29133284]
43. Connelly LM. Pilot studies. Medsurg Nurs 2008 Dec;17(6):411-412. [Medline: 19248407]
44. Morgado FF, Meireles JF, Neves CM, Amaral AC, Ferreira ME. Scale development: ten main limitations and recommendations to improve future research practices. Psicol Refl Crít 2017 Jan 25;30(1):1-20. [doi: 10.1186/s41155-016-0057-1]
45. Anthoine E, Moret L, Regnault A, Sébille V, Hardouin J. Sample size used to validate a scale: a review of publications on newly-developed patient reported outcomes measures. Health Qual Life Outcomes 2014 Dec 9;12:176 [FREE Full text] [doi: 10.1186/s12955-014-0176-2] [Medline: 25492701]
46. Polit DF, Beck CT. The content validity index: are you sure you know what's being reported? Critique and recommendations. Res Nurs Health 2006 Oct;29(5):489-497. [doi: 10.1002/nur.20147] [Medline: 16977646]
47. Ostrove JM, Adler NE, Kuppermann M, Washington AE. Objective and subjective assessments of socioeconomic status and their relationship to self-rated health in an ethnically diverse sample of pregnant women. Health Psychol 2000 Nov;19(6):613-618. [doi: 10.1037/0278-6133.19.6.613] [Medline: 11129365]
48. Adler NE, Epel ES, Castellazzo G, Ickovics JR. Relationship of subjective and objective social status with psychological and physiological functioning: preliminary data in healthy white women. Health Psychol 2000 Nov;19(6):586-592. [doi: 10.1037/0278-6133.19.6.586] [Medline: 11129362]
49. Norman C. eHealth literacy 2.0: problems and opportunities with an evolving concept. J Med Internet Res 2011 Dec 23;13(4):e125 [FREE Full text] [doi: 10.2196/jmir.2035] [Medline: 22193243]
50. Petric G, Atanasova S, Kamin T. Ill literates or illiterates? Investigating the eHealth literacy of users of online health communities. J Med Internet Res 2017 Oct 4;19(10):e331 [FREE Full text] [doi: 10.2196/jmir.7372] [Medline: 28978496]
51. Chang A, Schulz PJ. The measurements and an elaborated understanding of chinese ehealth literacy (C-eHEALS) in chronic patients in China. Int J Environ Res Public Health 2018 Jul 23;15(7):pii: E1553 [FREE Full text] [doi: 10.3390/ijerph15071553] [Medline: 30041420]
52. Sudbury-Riley L, FitzPatrick M, Schulz PJ. Exploring the measurement properties of the eHealth literacy scale (eHEALS) among baby boomers: a multinational test of measurement invariance. J Med Internet Res 2017 Feb 27;19(2):e53 [FREE Full text] [doi: 10.2196/jmir.5998] [Medline: 28242590]
53. Saffarzaden A. eScholarship. Reconceptualizing Health Literacy and the eHealth Literacy Scale (eHEALS): Evaluation of Psychometric Properties, Subdimensions, and Health-Related Internet Searching Behavior in Adult Outpatients Visiting a Tertiary Care Clinic URL: https://escholarship.org/uc/item/71x482hb [accessed 2019-10-09]
54. Soellner R, Huber S, Reder M. The concept of eHealth literacy and its measurement. J Media Psychol 2014;26(1):29-38. [doi: 10.1027/1864-1105/a000104]
55. Park H, Lee E. Self-reported eHealth literacy among undergraduate nursing students in South Korea: a pilot study. Nurse Educ Today 2015 Feb;35(2):408-413. [doi: 10.1016/j.nedt.2014.10.022] [Medline: 25466791]
56. Bazm S, Mirzae M, Fallahzadeh H, Bazm R. Validity and reliability of the Iranian version of eHealth literacy scale. J Commun Health Res 2016;5(2):121-130 [FREE Full text]
57. de Caro W, Corvo E, Marucci AR, Mitello L, Lancia L, Sansoni J. eHealth literacy scale: an nursing analisys and Italian validation. Stud Health Technol Inform 2016;225:949. [doi: 10.3233/978-1-61499-658-3-949] [Medline: 27332425]
58. Tubaishat A, Habiballah L. eHealth literacy among undergraduate nursing students. Nurse Educ Today 2016 Jul;42:47-52. [doi: 10.1016/j.nedt.2016.04.003] [Medline: 27237352]

59. Stellefson M, Paige SR, Tennant B, Alber JM, Chaney BH, Chaney D, et al. Reliability and validity of the telephone-based eHealth literacy scale among older adults: cross-sectional survey. J Med Internet Res 2017 Oct 26;19(10):e362 [FREE Full text] [doi: 10.2196/jmir.8481] [Medline: 29074471]
60. Richtering SS, Morris R, Soh S, Barker A, Bampi F, Neubeck L, et al. Examination of an eHealth literacy scale and a health literacy scale in a population with moderate to high cardiovascular risk: Rasch analyses. PLoS One 2017;12(4):e0175372 [FREE Full text] [doi: 10.1371/journal.pone.0175372] [Medline: 28448497]
61. Islam MM, Touray M, Yang H, Poly TN, Nguyen P, Li YJ, et al. E-Health literacy and health information seeking behavior among university students in Bangladesh. Stud Health Technol Inform 2017;245:122-125. [doi: 10.3233/978-1-61499-830-3-122] [Medline: 29295065]
62. Hyde LL, Boyes AW, Evans T, Mackenzie LJ, Sanson-Fisher R. Three-factor structure of the eHealth literacy scale among magnetic resonance imaging and computed tomography outpatients: a confirmatory factor analysis. JMIR Hum Factors 2018 Feb 19;5(1):e6 [FREE Full text] [doi: 10.2196/humanfactors.9039] [Medline: 29459356]
63. Paige SR, Miller MD, Krieger JL, Stellefson M, Cheong J. Electronic health literacy across the lifespan: measurement invariance study. J Med Internet Res 2018 Jul 9;20(7):e10434 [FREE Full text] [doi: 10.2196/10434] [Medline: 29986848]
64. Chung S, Park BK, Nahm E. The Korean eHealth literacy scale (K-eHEALS): reliability and validity testing in younger adults recruited online. J Med Internet Res 2018 Apr 20;20(4):e138 [FREE Full text] [doi: 10.2196/jmir.8759] [Medline: 29678800]
65. Chung S, Nahm E. Testing reliability and validity of the eHealth Literacy Scale (eHEALS) for older adults recruited online. Comput Inform Nurs 2015 Apr;33(4):150-156 [FREE Full text] [doi: 10.1097/CIN.0000000000000146] [Medline: 25783223]
66. Aponte J, Nokes KM. Electronic health literacy of older Hispanics with diabetes. Health Promot Int 2017 Jun 1;32(3):482-489. [doi: 10.1093/heapro/dav112] [Medline: 26681770]
67. Paige SR, Krieger JL, Stellefson M, Alber JM. eHealth literacy in chronic disease patients: an item response theory analysis of the eHealth literacy scale (eHEALS). Patient Educ Couns 2017 Feb;100(2):320-326 [FREE Full text] [doi: 10.1016/j.pec.2016.09.008] [Medline: 27658660]
68. Nguyen J, Moorhouse M, Curbow B, Christie J, Walsh-Childers K, Islam S. Construct validity of the eHealth literacy scale (eHEALS) among two adult populations: a Rasch analysis. JMIR Public Health Surveill 2016 May 20;2(1):e24 [FREE Full text] [doi: 10.2196/publichealth.4967] [Medline: 27244771]
69. Pérez GP, Almagro BJ, Gómez AH, Gómez JI. [Validation of the eHealth Literacy Scale (eHEALS) in Spanish University Students]. Rev Esp Salud Publica 2015;89(3):329-338 [FREE Full text] [doi: 10.4321/S1135-57272015000300010] [Medline: 26388346]
70. Mitsutake S, Shibata A, Ishii K, Oka K. Association of eHealth literacy with colorectal cancer knowledge and screening practice among internet users in Japan. J Med Internet Res 2012 Nov 13;14(6):e153 [FREE Full text] [doi: 10.2196/jmir.1927] [Medline: 23149453]
71. Asún RA, Rdz-Navarro K, Alvarado JM. The sirens' call in psychometrics: the invariance of IRT models. Theory Psychol 2017;27(3):389-406. [doi: 10.1177/0959354317706272]
72. Stewart AL, Thrasher AD, Goldberg J, Shea JA. A framework for understanding modifications to measures for diverse populations. J Aging Health 2012 Sep;24(6):992-1017 [FREE Full text] [doi: 10.1177/0898264312440321] [Medline: 22495768]
73. Kernisan LP, Sudore RL, Knight SJ. Information-seeking at a caregiving website: a qualitative analysis. J Med Internet Res 2010 Jul 28;12(3):e31 [FREE Full text] [doi: 10.2196/jmir.1548] [Medline: 20675292]
74. Lam L, Lam M. The use of information technology and mental health among older care-givers in Australia. Aging Ment Health 2009 Jul;13(4):557-562. [doi: 10.1080/13607860802607306] [Medline: 19629780]
75. Li H. Informal caregivers' use of the internet for caregiving information. Soc Work Health Care 2015;54(6):532-546. [doi: 10.1080/00981389.2015.1045577] [Medline: 26186424]
76. Anderson JG, Hundt E, Dean M, Keim-Malpass J, Lopez RP. 'The church of online support': examining the use of blogs among family caregivers of persons with dementia. J Fam Nurs 2017 Feb;23(1):34-54. [doi: 10.1177/1074840716681289] [Medline: 27920340]
77. Papastavrou E, Charalambous A, Tsangari H, Karayiannis G. The burdensome and depressive experience of caring: what cancer, schizophrenia, and Alzheimer's disease caregivers have in common. Cancer Nurs 2012;35(3):187-194. [doi: 10.1097/NCC.0b013e31822cb4a0] [Medline: 22538260]
78. Dimakopoulou E, Sakka P, Efthymiou A, Karpathiou N, Karydaki M. Internation Journal Of Caring Sciences. 2015. Evaluating the Needs of Dementia Patients' Caregivers in Greece: A Questionnaire Survey URL: http://internationaljournalofcaringsciences.org/docs/5_demakopoulou.pdf [accessed 2019-10-07]
79. Ruggiano N, Brown EL, Shaw S, Geldmacher D, Clarke P, Hristidis V, et al. The potential of information technology to navigate caregiving systems: perspectives from dementia caregivers. J Gerontol Soc Work 2019;62(4):432-450. [doi: 10.1080/01634372.2018.1546786] [Medline: 30422754]
80. Scharett E, Madathil KC, Lopes S, Rogers H, Agnisarman S, Narasimha S, et al. An investigation of information sought by caregivers of Alzheimer's patients on online peer-support groups. Cyberpsychol Behav Soc Netw 2017 Oct;20(10):640-657. [doi: 10.1089/cyber.2017.0274] [Medline: 29039697]

81. Jeong J, Kim Y, Chon M. Who is caring for the caregiver? The role of cybercoping for dementia caregivers. Health Commun 2018 Jan;33(1):5-13. [doi: 10.1080/10410236.2016.1242030] [Medline: 27901347]

Abbreviations

CFA: confirmatory factor analysis
CVI: content validity index
EFA: exploratory factor analysis
eHealth: electronic health
eHeals: eHealth Literacy Scale
eHeals-Carer: eHealth Literacy Scale for Carers of People With Chronic Diseases
I-CVI: Item-Level Content Validity Index
PCA: principal component analysis
S-CVI/Ave: Scale-Level Content Validity Index/Average
S-CVI/UA: Scale-Level Content Validity Index/Universal Agreement

Edited by G Eysenbach; submitted 14.10.18; peer-reviewed by S Paige, M Tremblay, S Atanasova, D Giroux, G Petric; comments to author 04.04.19; revised version received 28.05.19; accepted 31.08.19; published 28.11.19.

Please cite as:
Efthymiou A, Middleton N, Charalambous A, Papastavrou E
Adapting the eHealth Literacy Scale for Carers of People With Chronic Diseases (eHeals-Carer) in a Sample of Greek and Cypriot Carers of People With Dementia: Reliability and Validation Study
J Med Internet Res 2019;21(11):e12504
URL: http://www.jmir.org/2019/11/e12504/
doi:10.2196/12504
PMID:31778120

Original Paper

Adverse Events Due to Insomnia Drugs Reported in a Regulatory Database and Online Patient Reviews: Comparative Study

Jill S Borchert[1], PharmD; Bo Wang[2], BS; Muzaina Ramzanali[1], PharmD; Amy B Stein[3], PhD; Latha M Malaiyandi[4], PhD; Kirk E Dineley[4], PhD

[1]Chicago College of Pharmacy, Midwestern University, Downers Grove, IL, United States

[2]Chicago College of Osteopathic Medicine, Midwestern University, Downers Grove, IL, United States

[3]Office of Research and Sponsored Programs, Midwestern University, Glendale, AZ, United States

[4]College of Graduate Studies, Midwestern University, Downers Grove, IL, United States

Corresponding Author:
Kirk E Dineley, PhD
College of Graduate Studies
Midwestern University
555 31st Street
Downers Grove, IL, 60515
United States
Phone: 1 6309603907
Email: kdinel@midwestern.edu

Abstract

Background: Patient online drug reviews are a resource for other patients seeking information about the practical benefits and drawbacks of drug therapies. Patient reviews may also serve as a source of postmarketing safety data that are more user-friendly than regulatory databases. However, the reliability of online reviews has been questioned, because they do not undergo professional review and lack means of verification.

Objective: We evaluated online reviews of hypnotic medications, because they are commonly used and their therapeutic efficacy is particularly amenable to patient self-evaluation. Our primary objective was to compare the types and frequencies of adverse events reported to the Food and Drug Administration Adverse Event Reporting System (FAERS) with analogous information in patient reviews on the consumer health website Drugs.com. The secondary objectives were to describe patient reports of efficacy and adverse events and assess the influence of medication cost, effectiveness, and adverse events on user ratings of hypnotic medications.

Methods: Patient ratings and narratives were retrieved from 1407 reviews on Drugs.com between February 2007 and March 2018 for eszopiclone, ramelteon, suvorexant, zaleplon, and zolpidem. Reviews were coded to preferred terms in the Medical Dictionary for Regulatory Activities. These reviews were compared to 5916 cases in the FAERS database from January 2015 to September 2017.

Results: Similar adverse events were reported to both Drugs.com and FAERS. Both resources identified a lack of efficacy as a common complaint for all five drugs. Both resources revealed that amnesia commonly occurs with eszopiclone, zaleplon, and zolpidem, while nightmares commonly occur with suvorexant. Compared to FAERS, online reviews of zolpidem reported a much higher frequency of amnesia and partial sleep activities. User ratings were highest for zolpidem and lowest for suvorexant. Statistical analyses showed that patient ratings are influenced by considerations of efficacy and adverse events, while drug cost is unimportant.

Conclusions: For hypnotic medications, online patient reviews and FAERS emphasized similar adverse events. Online reviewers rated drugs based on perception of efficacy and adverse events. We conclude that online patient reviews of hypnotics are a valid source that can supplement traditional adverse event reporting systems.

(J Med Internet Res 2019;21(11):e13371) doi:10.2196/13371

KEYWORDS

drug safety; drug ineffective; postmarketing; pharmacovigilance; internet; pharmacoepidemiology; adverse effect; hypnotic; insomnia; patient-reported outcomes

Introduction

Postmarketing surveillance is coordinated by regulatory bodies that use passive collection systems to monitor the occurrence of drug toxicities. The Food and Drug Administration (FDA) Adverse Event Reporting System (FAERS) relies on drug manufacturers, health care professionals, and the patients themselves to report instances of adverse events in the form of an Individual Case Safety Report [1,2]. Although manufacturers are obligated to report adverse events, reporting by health care professionals and patients is voluntary. Consequently, a large amount of safety data never reach regulators [3-5], which underscores the need for additional mechanisms of pharmacovigilance.

Over the past decade, a rapidly expanding body of research has focused on the abundance of health data generated by patient online activities [6-10]. These include data collected from Web browser search habits [11,12], commentary that appears on social media such as Facebook and Twitter [13-16], and information revealed on internet forums [17,18]. The field of inquiry concerned with mining these resources for the purposes of improving public health has been referred to as "infodemiology" [19].

One type of Web datum enriched with drug safety information appears on websites such as WebMD, AskaPatient, and Drugs.com in the form of patient drug reviews. In a few sentences or a short paragraph, the patient shares his/her personal experience with efficacy, adverse events, and other issues related to the use of a given drug. Researchers have begun to explore how these reviews might improve pharmacovigilance. For example, online reviews have been used to show that emotional and behavioral effects are prominent considerations for users of antidepressants and antipsychotics [20,21]. Others found that online reviewers tend to describe less serious adverse events than those described by FAERS reports [22]. Web reviews can also help assess illicit drugs that fall outside of conventional trial evaluation [23], although evidence suggests that online commentary tends to minimize the dangers of habit-forming drugs [24]. Taken together, these studies suggest that online patient drug reviews can enhance traditional pharmacovigilance mechanisms by offering quick collection of information from large, diverse populations, followed by rapid dissemination in a form that is accessible to the lay community.

This study compares information contributed by patients to an online drug information website with adverse event data collected by the FDA. We focused on a particular set of FDA-approved hypnotics because they are common drugs used almost exclusively for the treatment of sleep disorders [25,26]. This minimized concerns over variable patient experience that arises when the same drugs are used for different therapeutic applications. In addition, hypnotics are well suited to patient evaluation, because the goal of therapy is simple, and most unwanted effects are easily identified. In our primary objective, we compared the frequencies of adverse events reported in FAERS with analogous information that appeared in patient reviews on the website Drugs.com. Secondary objectives were to describe patient reports of efficacy and adverse events and determine whether cost, effectiveness, or adverse events influenced user ratings of hypnotic medications.

Methods

Hypnotics

Five hypnotics were selected for this study. Eszopiclone, zaleplon, and zolpidem are a class of benzodiazepine receptor agonist known as Z-drugs and are the most commonly prescribed class of hypnotics [27]. Ramelteon is a melatonin receptor agonist that promotes sleep via activation of the melatonin 1 receptor subtype (MT1) [28]. Suvorexant promotes sleep by blocking OX1 and OX2 orexin receptors and is the first dual orexin receptor antagonist (DORA) approved for clinical use [29]. Other drugs used for insomnia (eg, benzodiazepines, trazodone, antihistamines, and melatonin) were excluded because those agents are frequently used for indications unrelated to sleep disorder or they are not FDA approved for the treatment of sleep disorders.

Online Reviews

A total of 1407 publicly available online drug reviews concerning either eszopiclone (n=239), ramelteon (n=72), suvorexant (n=324), zaleplon (n=82), or zolpidem (n=690), dated from February 2007 to March 2018, were retrieved from the website Drugs.com, a drug information platform for consumers and health care professionals. Drugs.com allows users to summarize their overall drug experience via anonymous text narratives and a numerical rating system, with 1 indicating not effective and 10 indicating most effective [30]. The total number of online reviews for a given drug is usually greater than the number of numerical ratings for that drug because not all reviewers chose to contribute a numerical summary rating.

Text narratives and numerical ratings from patient reviews were imported into Microsoft Excel (Microsoft Corporation, Redmond, Washington) for evaluation and analysis. A primary coder read each review to identify language used by the patient to convey an adverse event. Those keywords and phrases were then used to manually select low-level terms (LLTs) within MedDRA (Medical Dictionary for Regulatory Activities), version 18.0 [31]. MedDRA terminology is the international medical terminology developed under the auspices of the International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH). The corresponding preferred terms (PTs) were dictated by the LLT choice, consistent with MedDRA "Points to Consider" guidelines. Each review was also coded for mention of complaints of drug cost or insurance coverage. Microsoft Excel was then used to randomly select a subset of 166 cases that included at least 10% of the reviews for each of the five drugs. This subset was recoded by two secondary coders working independently and blinded to the LLT and PT selections of the primary coder. The percentage of reviews coded identically between the primary and the secondary coders was 74.7% and 78.3%, respectively. All three coders then met in person to discuss several recurring themes of disagreement apparent in the subset. For example, some reviews included the term "tolerance" to describe a drug effect that was diminished or lost with continued use ("I developed tolerance" or "I gained

tolerance"). This was variably coded with the LLT *drug effect decreased* or *drug tolerance increased*. The coders concluded that a clinician best assesses tolerance because it is an advanced concept that may include different pharmacokinetic, pharmacodynamic, and behavioral dimensions. In contrast, the patient is often best positioned to determine whether a hypnotic is still working or if the effect has waned. Therefore, narratives that mentioned tolerance were recoded as the LLT *drug effect decreased*. Reviews mentioning "depression" or "feeling depressed" were similarly problematic. The coders decided that the LLT *depression* refers to a formal, clinical diagnosis describing a set of symptoms persisting for a minimal length of time and that the mood changes implied in such narratives were more appropriately captured with the LLT *depressed mood*. Adjustments based on these and other term selections improved primary-secondary coder PT agreement to 79.6% and 87.3% of cases, within the subset. Those coding adjustments were then implemented across the entire data set. A small number of reviews mentioned or implied recent or concomitant use of other drugs, but we did not attempt to adjust for that in our data because those instances were rare and difficult to interpret. The primary and secondary coders completed MedDRA training workshops and webinars.

Food and Drug Administration Adverse Event Reporting System

For FAERS data, 11 quarterly reports (2015-Q1 through 2017-Q3) were downloaded from the FDA website [32]. Microsoft Access (Microsoft Corporation) was used to select 49,389 reports for eszopiclone, ramelteon, suvorexant, zaleplon, or zolpidem based on the "product_ai" field of the DRUG file. To limit the size and complexity of the data set, all reports with multiple drugs were excluded by eliminating cases with more than one "primaryid" entry; subsequently, only reports designating one of the five hypnotics as the primary suspect in the "role_cod" field were selected. These were matched back to the REAC, DEMO, and INDI files to retrieve the adverse event PTs, indications for therapy and demographics. For cases that appeared in multiple quarters due to report updating, we utilized the most recent version. This produced 5916 unique reports concerning either eszopiclone (n=196), ramelteon (n=103), suvorexant (n=4095), zaleplon (n=37), and zolpidem (n=1485) as the only drug reported. These were tabulated in Microsoft Excel for further analysis.

Statistical Analysis

Statistical analyses were performed using R statistical software (version 3.5.1; R Core Team, Vienna, Austria) using a two-sided significance level of .05. Statistical comparison of user drug ratings was performed using the nonparametric Kruskal-Wallis test, followed by the Dunn posthoc for all pairwise comparisons. To determine the relationship between user rating and complaints about efficacy or cost in the narrative, univariate logistic regression models were fit for each of the drugs separately. Logistic regression coefficients were reported as odds ratios with 95% CIs. The count of distinct adverse events recorded from the reviews (not including the PTs *drug ineffective* and *drug effect incomplete*) was analyzed using Poisson regression, with user rating as the explanatory variable. The exponential of the Poisson regression model coefficients was reported as the incidence rate estimates, and robust standard errors were calculated using the Delta method to control for mild violation of the distribution assumption that the variance equals the mean.

Ethics

The Institutional Review Board at Midwestern University - Downers Grove declared that this project does not qualify as human subjects research. The annotated datasets generated and analyzed during this study are available from the corresponding author on reasonable request.

Results

Adverse Event Reporting in Online Reviews Versus Food and Drug Administration Adverse Event Reporting System

The 10 most common MedDRA PT adverse events coded from online patient reviews are shown in Tables 1 and 2, expressed as a percentage of the total number of PTs recorded from all the reviews for each hypnotic. *Drug ineffective* was commonly reported for all five drugs (eszopiclone: 45/319, 14.1%; ramelteon: 24/104, 23.1%; suvorexant: 159/567, 28.0%; zaleplon: 22/82, 27%; and zolpidem: 33/958, 3.4%). Furthermore, partial or limited efficacy was noted in complaints that were coded as *drug effect incomplete* for ramelteon (12/104, 11.5%) and suvorexant (22/567, 3.9%). *Amnesia* was among the top 10 complaints for all three Z-drugs (eszopiclone: 8/319, 2.5%; zaleplon: 4/82, 5%; zolpidem: 161/958, 16.8%). Certain adverse events are notable for each of the different drugs. Zolpidem was frequently associated with complex partial sleep behaviors, including *abnormal sleep-related event* (90/958, 9.4%), *sleep-related eating disorder* (59/958, 6.2%), *somnambulism* (47/958, 4.9%), and *sleep talking* (34/958, 3.6%). Eszopiclone was commonly associated with *dysgeusia* (94/319, 29.5%). Suvorexant reviewers reported distressing parasomnias that included *nightmare* (54/567, 9.5%), *sleep paralysis* (26/567, 4.6%), and *abnormal dreams* (25/567, 4.4%).

Tables 3 and 4 show the top 10 most common PTs in FAERS reports, expressed as the percentage of the total number of PTs for each drug. *Drug ineffective* emerged as the most common PT recorded in FAERS for each of the five drugs (eszopiclone: 78/458, 17.0%; ramelteon: 30/208, 14.4%; suvorexant: 1108/6171, 18.0%; zaleplon: 18/77, 23%; zolpidem: 499/3448, 14.5%). Other PTs common for the individual drugs include *amnesia* (51/3448, 1.5%) and *somnambulism* (75/3448, 2.2%) with zolpidem; *dysgeusia* (22/458, 4.8%) and *product substitution issue* (30/458, 6.6%) with eszopiclone; and *nightmare* (422/6171, 6.8%), *abnormal dreams* (382/6171, 6.2%), and *sleep paralysis* (124/6171, 2.0%) with suvorexant.

Table 1. Preferred terms manually coded from online patient reviews of Z-drugs.

#	Eszopiclone (n=319)		Zaleplon (n=82)		Zolpidem (n=958)	
	Preferred term	n (%)	Preferred term	n (%)	Preferred term	n (%)
1	Dysgeusia	94 (29.5)	Drug ineffective	22 (27)	Amnesia	161 (16.8)
2	Drug ineffective	45 (14.1)	Insomnia	5 (6)	Abnormal sleep-related event	90 (9.4)
3	Drug effect decreased	17 (5.3)	Somnolence	5 (6)	Sleep-related eating disorder	59 (6.2)
4	Product substitution issue	17 (5.3)	Amnesia	4 (5)	Drug effect decreased	51 (5.3)
5	Insomnia	10 (3.1)	Abnormal sleep-related event	3 (4)	Somnambulism	47 (4.9)
6	Depressed mood	9 (2.8)	Anxiety	3 (4)	Sleep talking	34 (3.6)
7	Somnolence	9 (2.8)	Hallucination	3 (4)	Drug ineffective	33 (3.4)
8	Amnesia	8 (2.5)	Headache	3 (4)	Drug dependence	32 (3.3)
9	Anxiety	7 (2.2)	Restless legs syndrome	3 (4)	Somnolence	28 (2.9)
10	Headache	6 (1.9)	Abdominal pain upper	2 (2)	Hallucination	27 (2.8)

Table 2. Preferred terms manually coded from online patient reviews of ramelteon and suvorexant.

#	Ramelteon (n=104)		Suvorexant (n=567)	
	Preferred term	n (%)	Preferred term	n (%)
1	Drug ineffective	24 (23.1)	Drug ineffective	159 (28.0)
2	Drug effect incomplete	12 (11.5)	Nightmare	54 (9.5)
3	Somnolence	7 (6.7)	Headache	27 (4.8)
4	Dizziness	5 (4.8)	Somnolence	27 (4.8)
5	Insomnia	5 (4.8)	Sleep paralysis	26 (4.6)
6	Abnormal dreams	4 (3.9)	Abnormal dreams	25 (4.4)
7	Depressed mood	4 (3.9)	Drug effect incomplete	22 (3.9)
8	Feeling abnormal	4 (3.9)	Feeling abnormal	20 (3.5)
9	Anxiety	3 (2.9)	Insomnia	19 (3.4)
10	Headache	3 (2.9)	Hangover	17 (3.0)

Table 3. Preferred terms retrieved from FAERS reports for Z-drugs.

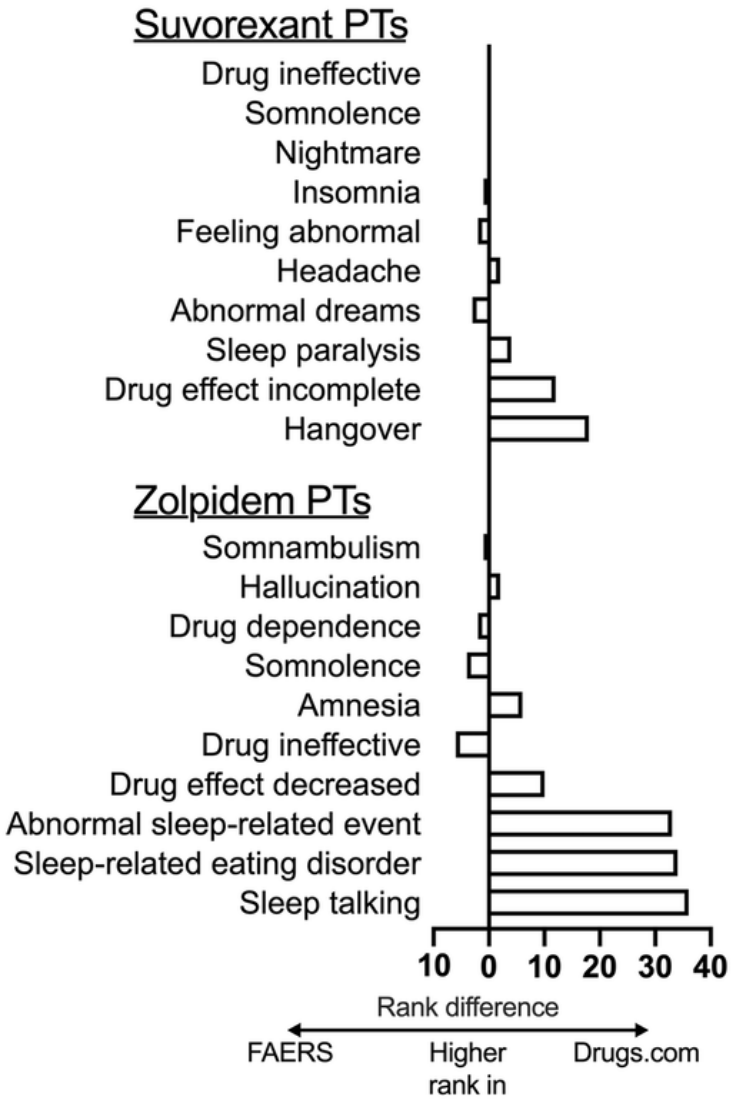
#	Eszopiclone (n=458)		Zaleplon (n=77)		Zolpidem (n=3448)	
	Preferred term	n (%)	Preferred term	n (%)	Preferred term	n (%)
1	Drug ineffective	78 (17.0)	Drug ineffective	18 (23)	Drug ineffective	499 (14.5)
2	Insomnia	33 (7.2)	Drug effect incomplete	3 (4)	Insomnia	146 (4.2)
3	Product substitution issue	30 (6.6)	Insomnia	3 (4)	Product substitution issue	96 (2.8)
4	Dysgeusia	22 (4.8)	Product quality issue	3 (4)	Somnambulism	75 (2.2)
5	Product quality issue	16 (3.5)	Product substitution issue	3 (4)	Somnolence	59 (1.7)
6	Drug effect decreased	14 (3.1)	Abnormal behavior	2 (3)	Drug dependence	57 (1.7)
7	Delirium	12 (2.6)	Drug hypersensitivity	2 (3)	Amnesia	51 (1.5)
8	Headache	7 (1.5)	Malaise	2 (3)	Road traffic accident	48 (1.4)
9	Nausea	7 (1.5)	Nausea	2 (3)	Toxicity to various agents	47 (1.4)
10	Sleep disorder	7 (1.5)	Nightmare	2 (3)	Overdose	46 (1.3)

Table 4. Preferred terms retrieved from FAERS reports for ramelteon and suvorexant.

#	Ramelteon (n=208)		Suvorexant (n=6171)	
	Preferred term	n (%)	Preferred term	n (%)
1	Drug ineffective	30 (14.4)	Drug ineffective	1108 (18.0)
2	No adverse event	17 (8.2)	Nightmare	422 (6.8)
3	Intentional overdose	14 (6.7)	Abnormal dreams	382 (6.2)
4	Toxicity to various agents	13 (6.3)	Somnolence	256 (4.2)
5	Somnolence	8 (3.9)	Headache	199 (3.2)
6	Suicide attempt	8 (3.9)	Feeling abnormal	189 (3.1)
7	Drug prescribing error	7 (3.4)	Hallucination	174 (2.8)
8	Dizziness	6 (2.9)	Insomnia	136 (2.2)
9	Middle insomnia	6 (2.9)	Sleep paralysis	124 (2.0)
10	Drug administered to patient of inappropriate age	5 (2.4)	Adverse event	109 (1.8)

To graphically summarize areas of agreement and disagreement between Drugs.com and FAERS, we plotted the difference in rank positions of the most frequent PTs in Drugs.com from their position in FAERS for suvorexant and zolpidem (Figure 1). In this representation, a small bar indicates PTs that were ranked similarly in Drugs.com and FAERS, while a larger bar shows differing ranks between the two lists. For zolpidem, for example, *sleep talking* ranked as the 6th most common PT in Drugs.com, but only the 42nd most common in FAERS; thus, the rank difference is 36. There were insufficient data from one or both sources to provide similar representations for the other three drugs studied.

Figure 1. Difference in rank position for the 10 most common PTs in Drugs.com vs their rank in FAERS, for suvorexant and zolpidem. The rank of the Drugs.com PT was subtracted from its corresponding rank in FAERS. Bars extend to the right for PTs that held higher rank position in Drugs.com than in FAERS. For PTs that held higher rank position in FAERS compared to Drugs.com, bars extend to the left. PT: preferred term; FAERS: Food and Drug Administration Adverse Event Reporting System.



Online Review User Ratings

We next considered the Drugs.com online reviews in more detail. Figure 2A depicts the accumulation of reviews for each drug on a monthly basis for the period of February 2007 through March 2018. As of March 2018, zolpidem had the most reviews (n=690), followed by suvorexant (n=324), eszopiclone (n=239), zaleplon (n=82), and ramelteon (n=72). Figure 2B depicts a running monthly average of the numerical 1-10 user ratings. Zolpidem had an average rating of 7.30, followed by the other two Z-drugs, eszopiclone (6.20) and zaleplon (5.69). The nonbenzodiazepine receptor drugs ramelteon and suvorexant were rated 4.63 and 3.65, respectively. Numerical ratings were tabulated in frequency histograms (Figure 2C), which yielded bimodal distributions for each of the five drugs. These data show that a high percentage of reviewers assigned suvorexant and ramelteon the lowest possible score of "1" (154/290, 53.1% and 27/65, 41.5%, respectively). In comparison, only 10.4% (66/637) of zolpidem reviewers rated it as "1," while 31.6% (201/637) rated it "10." Analysis of user ratings using Kruskal-Wallis test followed by the Dunn post hoc test found that zolpidem was rated significantly higher than the other four drugs, while suvorexant was rated significantly lower than all the Z-drugs (Table 5).

We then explored factors that might contribute to high or low user ratings. Focusing specifically on comments of poor efficacy, we found a statistically significant association between low user ratings and reviews that were coded for the PTs *drug ineffective* or *drug effect incomplete* for all five drugs (Figure 3). The odds ratios represent the increased odds of an ineffective complaint relative to a unit decrease in user rating. The number of reviews that were coded for *drug ineffective* or *drug effect incomplete* was 44/239 (18.4%) for eszopiclone, 25/72 (35%) for ramelteon, 165/324 (50.9%) for suvorexant, 17/82 (20.7%) for zaleplon, and 34/690 (4.9%) for zolpidem. In addition, a Poisson regression analysis found a significant inverse correlation between the numerical rating and the number of distinct PTs captured from a review for all five drugs (Table 6).

Figure 2. Statistics for consumer reviews of hypnotics on the website Drugs.com. (A) Accumulation of consumer reviews per month. Running total number of reviews are depicted for eszopiclone, ramelteon, suvorexant, zaleplon, and zolpidem. (B) Running monthly average, weighted by frequency, based on reviewers' numerical summary ranking (1-10), with 1 indicating a very poor experience and 10 indicating a very positive experience. (C) Percent frequency histograms for numerical (1-10) user ratings of insomnia drugs from drugs.com. Data are inclusive of reviews posted between February 2007 and March 2018.

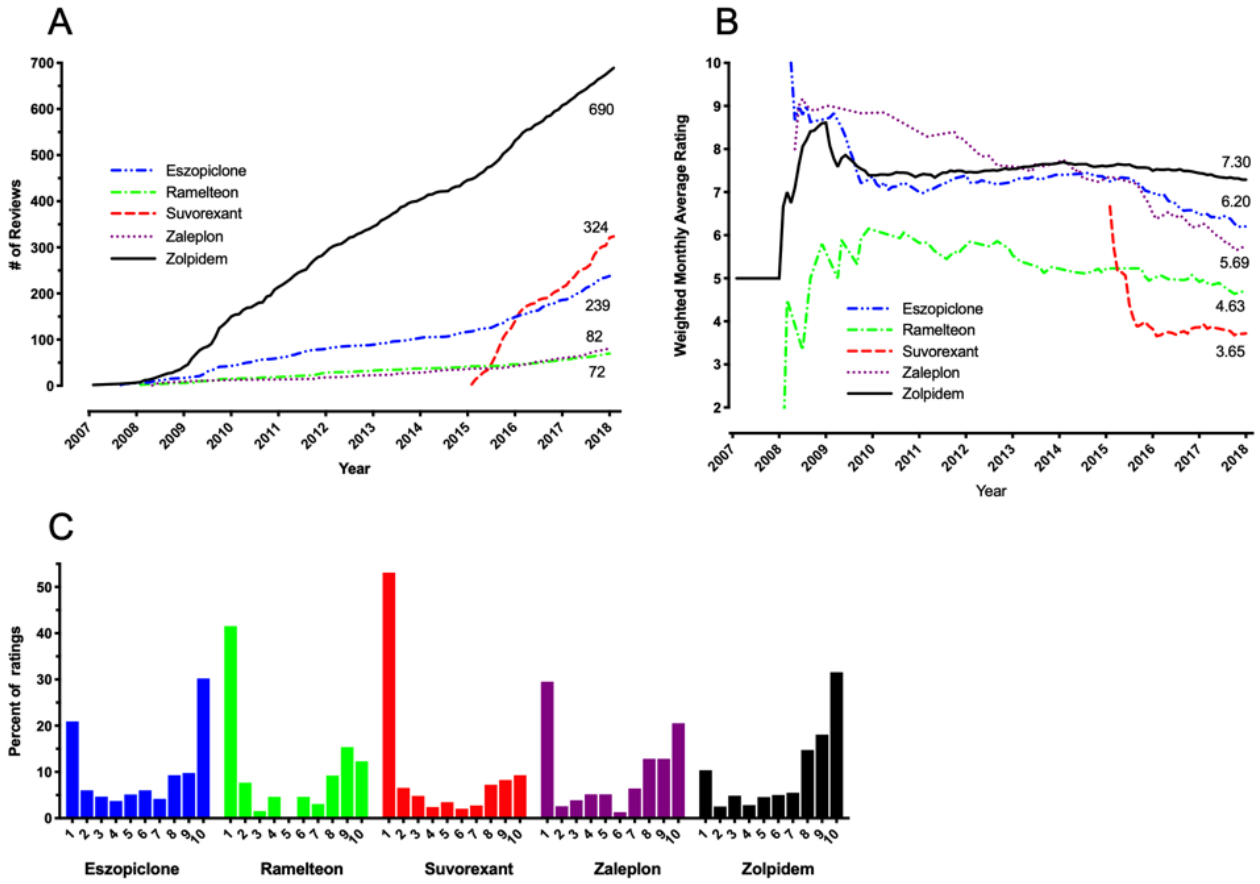


Table 5. Comparisons of hypnotic user ratings in Drugs.com with Kruskal-Wallis followed by Dunn post hoc.

Drug - comparator (median user rating)	Adjusted *P* value
Ramelteon (3) - eszopiclone (7)	.003
Suvorexant (1) - eszopiclone (7)	<.001
Suvorexant (1) - ramelteon (3)	.10
Zaleplon (7) - eszopiclone (7)	.09
Zaleplon (7) - ramelteon (3)	.11
Zaleplon (7) - suvorexant (1)	<.001
Zolpidem (8) - eszopiclone (7)	.003
Zolpidem (8) - ramelteon (3)	<.001
Zolpidem (8) - suvorexant (1)	<.001
Zolpidem (8) - zaleplon (7)	<.001

Figure 3. Odds ratio plot of the univariate logistic regression models to assess the relationship between user ratings and patient complaints of poor efficacy, that is, reviews that were coded for *drug ineffective* or *drug effect incomplete*. Odds ratios and 95% CIs are depicted by the circle and whiskers, respectively.

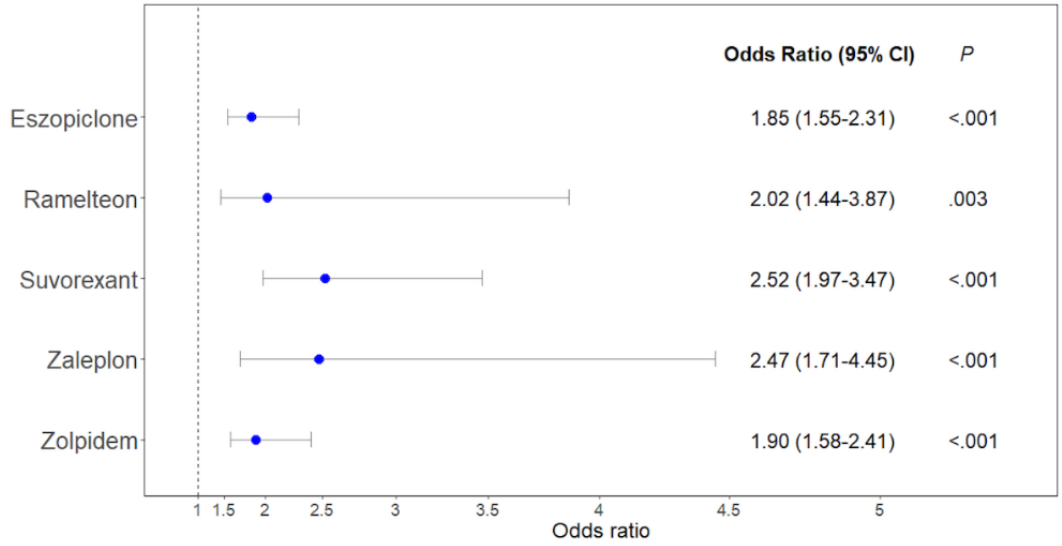


Table 6. Estimated incidence rate ratios for the effect of user ratings on the number of adverse events coded from patient reviews. The expected count of PTs is multiplied by a factor of the estimated incidence rate ratios when the user rating increases by one unit.

Drug	Incidence rate ratios (95% CI)	*P* value
Eszopiclone	0.89 (0.85-0.92)	<.001
Ramelteon	0.90 (0.83-0.97)	.007
Suvorexant	0.90 (0.86-0.95)	<.001
Zaleplon	0.88 (0.79-0.98)	<.001
Zolpidem	0.88 (0.86-0.90)	<.001

The estimated effect sizes correspond to a percent change in the incidence rate of PTs by 11% in eszopiclone, 10% in ramelteon, 10% in suvorexant, 12% in zaleplon, and 12% in zolpidem for a one-score decrease in the user rating. Considering eszopiclone, for example, this suggests that the expected count of PTs decreased by a factor of 0.89 when the user rating increased by one unit.

Users also mentioned cost in the reviews, most commonly with suvorexant (75/324, 23.1%) and eszopiclone (37/239, 15.5%). Univariate logistic regression showed that the tendency to mention cost did not predict lower numerical ratings for any of the drugs (Multimedia Appendix 1). For eszopiclone, however, each one-unit increase in user rating increased the odds of a comment about cost by 1.24, suggesting that cost commentary may be correlated with a *higher* eszopiclone rating. Zaleplon was not included in this analysis due to insufficient data, with only 1 of 82 reviews mentioning cost.

Drugs.com does not collect demographic data for individual reviewers, but they compile top-level information on user age and country of residence, which was provided to us after sending an email request to their user support (Multimedia Appendix 2). These data show that more than half (54%) of the reviews are contributed by patients aged ≤44 years and that most reviewers are female (57%) and from the United States (70%).

We also extracted from FAERS patient demographics data on indication of therapy and reporter information, which are summarized in Multimedia Appendix 3. A plurality of reports concerned female patients (2935/5916, 49.6%), and the most common indications were sleep disorders (3735/5916, 63.1%). A substantial fraction of the cases did not report gender (1122/5916, 19%), age (4054/5916, 68.5%), or indication (2102/5916, 35.5%). Most cases (4618/5916, 78.1%) were reported by the consumers themselves, while nearly all of the remainder were reported by health care professionals. More than 9/10 FAERS cases were contributed by reporters in the United States (5383/5916, 91%). Patient age was normally distributed around a mean of 56.9 (SD 18.2) years (Multimedia Appendix 4).

Discussion

Principal Findings

In general, we found that adverse event data interpreted from patient online reviews were consistent with the adverse event data in FAERS. Both resources show that poor efficacy is common with all five drugs and that amnesia is frequent with Z-drugs. A variety of adverse events particular to the individual drugs were also evident in both, for example, *nightmare* with suvorexant, *dysgeusia* and *product substitution issue* with eszopiclone, and *dizziness* with ramelteon.

The two sources showed good agreement in zolpidem adverse events (*drug ineffective*, *somnambulism*, *somnolence*, *drug dependence*, and *amnesia* were among the top 10 most frequent PTs in both), but the online reviews contained a higher rate of partial sleep activities, such as walking, talking, and eating while semiconscious (for sample reviews, see Multimedia Appendix 5). Reviewers often did not remember those events, which in turn contributed to a high incidence of *amnesia* in the online data versus FAERS data (16.8% vs 1.5%). This discrepancy between the two sources may be explained by the fact that the propensity for zolpidem-induced partial sleep activities is now so well known that health professionals see little value in reporting it, which would lead to an underestimation of those events in recent FAERS quarterlies. The opposite trend may be occurring in the online review data, where many reviewers seemed eager to share those stories. Such perspectives might be expected from a younger population that is more likely to participate in online activity [22,33] and appears to be more heavily represented in Drugs.com compared to FAERS. Indeed, patient age is one of the notable differences between the two resources: The average age in FAERS was nearly 60 years, while over half of the Drugs.com user population is aged ≤44 years. Regardless, the high numerical ratings for zolpidem (Table 5) suggest that online reviewers may be trivializing the very serious dangers of Z-drug impairment [34], which recently prompted an FDA-mandated boxed warning for all three Z-drugs [35]. Other studies have found a general disregard of serious safety issues among the online community. For example, official safety warning announcements regarding zolpidem registered on social media in only limited and transient fashion [36], and online reviewers minimized the dangers associated with sibutramine, a weight-loss drug that was ultimately withdrawn for safety reasons [22]. A related issue was described by Adusumalli et al [24], who noted that patients may tend to rate addictive drugs more highly than alternatives that are less addictive but may be equally effective [24]. This suggests that patient opinion of zolpidem and other Z-drugs might be positively correlated with their habit-forming tendencies.

For suvorexant, we found strong agreement between the two sources. Eight adverse event PTs were among the top 10 most frequent in both sources: *drug ineffective*, *nightmare*, *somnolence*, *headache*, *sleep paralysis*, *abnormal dreams*, *feeling abnormal*, and *insomnia*. The high incidence of *drug ineffective* is in agreement with the reported incidence in another recent study [37], which found that suvorexant is one of the most common drugs associated with this complaint in the FAERS database. Parasomnias were also frequent, especially nightmares, which deserve special comment because they seem to be rather more intense than a typical “scary” dream, with numerous patients describing them in exceptionally vivid and terrifying themes (Multimedia Appendix 5). Here, it is important to note that nightmare was not mentioned as a potential adverse event in the suvorexant prescribing information [38] nor is it discussed in an exhaustive review of the discovery and clinical development of suvorexant [29]. This discrepancy might be explained if the premarket trials detected only mildly disturbing dreams that were coded not as *nightmare* but as *abnormal dreams*—an adverse event that is explicitly described in the suvorexant prescribing information. However, our findings indicate that conflating *nightmare* with *abnormal dreams* does not provide a true picture of the user experience with this drug. This observation adds to previous studies, which also showed that patient-contributed online data can help capture adverse events that were not described in trial data evaluated by the FDA [39,40].

Factors Influencing Summary Drug Rating

For all five drugs, negative ratings significantly increased the odds of the PT drug ineffectiveness and the frequency of distinct PTs coded from the text of the corresponding review. A substantial number of online reviews for suvorexant, ramelteon, and eszopiclone explicitly mentioned cost or insurance coverage, a concern that is likely to be more prominent with reviewers in the United States compared to patients from countries with national health care programs. Unexpectedly, we found that commentary about affordability did not predict low user ratings for any of the drugs, and eszopiclone reviewers who commented about cost were actually more likely to assign a *higher* numerical rating. We think this seeming paradox is explained by the advent of generic formulations of eszopiclone in the United States during the review period, which led to complaints of *product substitution issue* from reviewers who perceived less benefit from the generic formulations (Multimedia Appendix 5). Accordingly, that PT did not appear in the eszopiclone reviews until September 2014. Around the same time, the previously stably running average began a gradual downward trend (Figure 3B).

To put the patient ratings in perspective, we considered several unrelated medications noted to cause distressing adverse events. The antipsychotic olanzapine causes significant weight gain and metabolic disorder but has a 6.7 average rating on Drugs.com [41]. Methotrexate, used frequently for rheumatoid arthritis, is associated with gastrointestinal upset, painful mouth ulcers, and fatigue, but has a rating of 7.0 [42]. The antibiotic clindamycin is rated 5.9, despite causing diarrhea in a high percentage of patients [43]. This brief sampling suggests that even drugs with onerous adverse events can earn a positive rating, as long as patients perceive real benefit. The exceedingly low ratings of suvorexant and ramelteon were even more remarkable when viewed in this context. In the larger view, our observations suggest that online patient evaluations of hypnotics are guided by fundamental considerations of drug efficacy and tolerability.

Patient Perspective and Practical Utility of Online Reviews

A patient-reported outcome is a patient’s assessment of their own response to therapeutic intervention that is not reinterpreted or filtered by a clinician [44,45]. Because patients tend to evaluate their response to therapy in a holistic sense, patient-reported outcomes are useful evaluations of functional status and quality of life. These patient-reported outcomes are collected with validated instruments during drug development trials, but after a drug is marketed, there is a continued need to capture the patient’s opinion in the assessment of therapeutic response. Online reviews seem well suited for recording the patient perspective, especially for drugs like hypnotics, where the balance of efficacy versus adverse events relates directly to

the functional benefits that accrue with therapeutic success, that is, improved sleep. Furthermore, hypnotics should be reconsidered with regard to the patient's perspective because guidance on examining patient-reported outcomes [46,47] was published after trials of several important hypnotic medications were completed [48]. In addition, clinical trial participants differ from real-world patients [49]; therefore, even hypnotics with historical patient-reported outcomes can benefit from continued surveillance of the patient perspective in the form of online reviews.

In the most practical terms, online reviews can serve as an initial source of information for patients seeking a mix of views on the potential advantages and disadvantages of a given drug therapy. Compared to professional resources with similar information, online reviews are written with language and tone that resonate with the average patient [21,50]. The recently introduced FAERS dashboard can also educate consumers about drug adverse events [51]; however, interpretation of those data is still quite challenging for untrained people. Regardless of the resource, health care professionals should be ready to correct a number of common misconceptions. Patients must understand that a report of an adverse event does not establish causality, nor can individual probability of experiencing an adverse event be inferred from adverse event reports. Such a discussion can help the clinician develop an optimal plan of care that fully considers patient concerns about adverse events and expectations of effectiveness.

Limitations

The most obvious limitation of this study is the suspect validity of patient online data. We observed no reason to doubt the sincerity of these reviews, but the general vulnerability of patient reviews to misinformation and manipulation must be kept in mind, as other investigators have cautioned [52]. The subjective nature of patient self-evaluations may also be an issue, although it should be noted that self-reported data are standard for clinical sleep studies [48], which often rely on the subjects themselves to record their data and observations in a daily "sleep diary" [53].

Inherent differences in size and complexity between the two data sources made it difficult to compare the data pertaining to the same time period. We needed all the online reviews available since Drugs.com introduced this feature (in ~2008); otherwise, those data sets would be too small. In contrast, even just a few years of FAERS data can become overwhelmingly large, and FAERS conventions undergo periodic changes to MedDRA coding practices and formatting of the quarterly files. In view of those complications, we opted for a simple solution that limited our FAERS survey to 11 quarters, following the approval of suvorexant in August 2014. This compromise facilitated a comparison of two sets of concurrent suvorexant data, the agent with the most limited clinical experience. Perhaps, this approach contributed to the high degree of agreement between FAERS and Drugs.com for suvorexant; if so, it probably also explains some of the disagreement for the other drugs, where the data timelines overlap less.

Other limitations concern the sparse nature of the online data compared to reports collected by regulatory agencies [54]. Although Drugs.com collects limited data on their user population, individuals rarely volunteered their age, gender, or medication dosage in the text of their reviews. Furthermore, it is often impossible to know what additional drugs they may have taken concomitantly, which could be the primary source of the adverse event. This consideration is especially relevant to these hypnotics, because all are subject to hepatic metabolism and attendant drug-drug interactions. Because follow-up clarification of case information is not possible with anonymous online testimonials lacking a verifiable patient, these limitations are fixed [55]. In contrast, we noted that many FAERS reports also lacked basic data including age, gender, and indication. Most revealing is the fact that more than three quarters of the FAERS cases (4618/5916, 78.1%) were reported by consumers, which means that these FAERS reports apparently share the same untrained origins as the online reviews, regardless of subsequent coding and processing by professionals.

Conclusions

Our work adds to a growing body of literature that has explored the utility of online reviews. Patient numerical ratings of hypnotic medications were influenced by perceptions of efficacy and adverse events, but not by cost. Patients rated zolpidem the highest, while ramelteon and suvorexant were held in relatively poor regard. Patient online reviews emphasized many of the same hypnotic adverse events that are reported to the FDA, most notably lack of efficacy. Future research is needed to determine how online reviews may help collect patient-reported outcomes in the postmarketing realm. Because our results show that online reviews are guided primarily by considerations of drug efficacy and tolerability and match well with adverse event data reported to FAERS, we conclude that patient online reviews offer a valuable supplement to traditional adverse event reporting systems.

Acknowledgments

We are grateful to Michelle Sloan for help with data analysis, and Ameir Barakat and Peter Lamar for proofreading. Midwestern University had no involvement in study design; collection, analysis, and interpretation of data; writing of the report; or the decision to submit the article for publication. MedDRA trademark is registered by the IFPMA on behalf of the ICH.

Authors' Contributions

KD took primary responsibility for conducting this study. All authors contributed to the conception and study design, and all authors participated in data collection, analyses, and interpretation. KD and JSB drafted the manuscript with support from LMM and ABS. All authors contributed to revisions of the manuscript and approved the final version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Estimated odd ratios for the relationships between cost complaints and user ratings.
[PDF File (Adobe PDF File), 43 KB - jmir_v21i11e13371_app1.pdf]

Multimedia Appendix 2
Top-level user demographics for drugs.com, as provided by drugs.com user support.
[PDF File (Adobe PDF File), 35 KB - jmir_v21i11e13371_app2.pdf]

Multimedia Appendix 3
Demographics of 5916 patient reports for five insomnia drugs in the Food and Drug Administration Adverse Event Reporting System. Unkn/NA: Unknown or not available; Sleep dis: Sleep disorders.
[PDF File (Adobe PDF File), 37 KB - jmir_v21i11e13371_app3.pdf]

Multimedia Appendix 4
Histogram of patient age data in Food and Drug Administration Adverse Event Reporting System reports (mean 56.9 years, SD 18.2 years; range 0-97 years). Age data were available for 31.5% (1862/5916) of reports. Cases recorded in nonyear units were converted to years. For cases recorded in decade units, the midpoint year was used, eg, “7 DEC” was considered 65 years of age.
[PDF File (Adobe PDF File), 51 KB - jmir_v21i11e13371_app4.pdf]

Multimedia Appendix 5
Example reviews from Drugs.com for eszopiclone, suvorexant, and zolpidem.
[PDF File (Adobe PDF File), 42 KB - jmir_v21i11e13371_app5.pdf]

References

1. Ahmad SR. Adverse drug event monitoring at the Food and Drug Administration. J Gen Intern Med 2003 Jan;18(1):57-60 [FREE Full text] [Medline: 12534765]
2. FDA: U.S. Food and Drug Administration. Individual Case Safety Reports. URL: https://www.fda.gov/industry/fda-resources-data-standards/individual-case-safety-reports [accessed 2019-08-11]
3. Chaplin S. The Yellow Card scheme - why are GPs under-reporting? Prescriber 2006 Aug 05;17(15):18-22. [doi: 10.1002/psb.405]
4. van der Hooft CS, Sturkenboom MCJM, van Grootheest K, Kingma HJ, Stricker B. Adverse drug reaction-related hospitalisations: a nationwide study in The Netherlands. Drug Saf 2006;29(2):161-168. [doi: 10.2165/00002018-200629020-00006] [Medline: 16454543]
5. Lopez-Gonzalez E, Herdeiro MT, Figueiras A. Determinants of under-reporting of adverse drug reactions: a systematic review. Drug Saf 2009;32(1):19-31. [Medline: 19132802]
6. Sloane R, Osanlou O, Lewis D, Bollegala D, Maskell S, Pirmohamed M. Social media and pharmacovigilance: A review of the opportunities and challenges. Br J Clin Pharmacol 2015 Oct;80(4):910-920 [FREE Full text] [doi: 10.1111/bcp.12717] [Medline: 26147850]
7. Sarker A, Ginn R, Nikfarjam A, O'Connor K, Smith K, Jayaraman S, et al. Utilizing social media data for pharmacovigilance: A review. J Biomed Inform 2015 Apr;54:202-212 [FREE Full text] [doi: 10.1016/j.jbi.2015.02.004] [Medline: 25720841]
8. Omar I, Harris E. The Use of Social Media in ADR Monitoring and Reporting. J Pharmacovigilance 2016;04(06). [doi: 10.4172/2329-6887.1000223]
9. Lardon J, Abdellaoui R, Bellet F, Asfari H, Souvignet J, Texier N, et al. Adverse Drug Reaction Identification and Extraction in Social Media: A Scoping Review. J Med Internet Res 2015 Jul 10;17(7):e171 [FREE Full text] [doi: 10.2196/jmir.4304] [Medline: 26163365]
10. Khare R, Good BM, Leaman R, Su AI, Lu Z. Crowdsourcing in biomedicine: challenges and opportunities. Brief Bioinform 2016 Jan;17(1):23-32. [doi: 10.1093/bib/bbv021] [Medline: 25888696]
11. White RW, Tatonetti NP, Shah NH, Altman RB, Horvitz E. Web-scale pharmacovigilance: listening to signals from the crowd. J Am Med Inform Assoc 2013 May 1;20(3):404-408 [FREE Full text] [doi: 10.1136/amiajnl-2012-001482] [Medline: 23467469]
12. Pierce CE, Bouri K, Pamer C, Proestel S, Rodriguez HW, Van Le H, et al. Evaluation of Facebook and Twitter Monitoring to Detect Safety Signals for Medical Products: An Analysis of Recent FDA Safety Alerts. Drug Saf 2017 Apr;40(4):317-331 [FREE Full text] [doi: 10.1007/s40264-016-0491-0] [Medline: 28044249]

13. Comfort S, Perera S, Hudson Z, Dorrell D, Meireis S, Nagarajan M, et al. Sorting Through the Safety Data Haystack: Using Machine Learning to Identify Individual Case Safety Reports in Social-Digital Media. Drug Saf 2018 Jun;41(6):579-590 [FREE Full text] [doi: 10.1007/s40264-018-0641-7] [Medline: 29446035]
14. Bollegala D, Maskell S, Sloane R, Hajne J, Pirmohamed M. Causality Patterns for Detecting Adverse Drug Reactions From Social Media: Text Mining Approach. JMIR Public Health Surveill 2018 May 09;4(2):e51 [FREE Full text] [doi: 10.2196/publichealth.8214] [Medline: 29743155]
15. Smith K, Golder S, Sarker A, Loke Y, O'Connor K, Gonzalez-Hernandez G. Methods to Compare Adverse Events in Twitter to FAERS, Drug Information Databases, and Systematic Reviews: Proof of Concept with Adalimumab. Drug Saf 2018 Dec;41(12):1397-1410 [FREE Full text] [doi: 10.1007/s40264-018-0707-6] [Medline: 30167992]
16. Freifeld CC, Brownstein JS, Menone CM, Bao W, Filice R, Kass-Hout T, et al. Digital drug safety surveillance: monitoring pharmaceutical products in twitter. Drug Saf 2014 May;37(5):343-350 [FREE Full text] [doi: 10.1007/s40264-014-0155-x] [Medline: 24777653]
17. Matsuda S, Aoki K, Tomizawa S, Sone M, Tanaka R, Kuriki H, et al. Analysis of Patient Narratives in Disease Blogs on the Internet: An Exploratory Study of Social Pharmacovigilance. JMIR Public Health Surveill 2017 Feb 24;3(1):e10 [FREE Full text] [doi: 10.2196/publichealth.6872] [Medline: 28235749]
18. Chen JH, Altman RB. Data-Mining Electronic Medical Records for Clinical Order Recommendations: Wisdom of the Crowd or Tyranny of the Mob? AMIA Jt Summits Transl Sci Proc 2015;2015:435-439 [FREE Full text] [Medline: 26306281]
19. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. J Med Internet Res 2009;11(1):e11 [FREE Full text] [doi: 10.2196/jmir.1157] [Medline: 19329408]
20. Hughes S, Lacasse J, Fuller RR, Spaulding-Givens J. Adverse effects and treatment satisfaction among online users of four antidepressants. Psychiatry Res 2017 Sep;255:78-86. [doi: 10.1016/j.psychres.2017.05.021] [Medline: 28531820]
21. Moncrieff J, Cohen D, Mason JP. The subjective experience of taking antipsychotic medication: a content analysis of Internet data. Acta Psychiatr Scand 2009 Aug;120(2):102-111. [doi: 10.1111/j.1600-0447.2009.01356.x] [Medline: 19222405]
22. Duh MS, Cremieux P, Audenrode MV, Vekeman F, Karner P, Zhang H, et al. Can social media data lead to earlier detection of drug-related adverse events? Pharmacoepidemiol Drug Saf 2016 Dec;25(12):1425-1433 [FREE Full text] [doi: 10.1002/pds.4090] [Medline: 27601271]
23. Swogger MT, Hart E, Erowid F, Erowid E, Trabold N, Yee K, et al. Experiences of Kratom Users: A Qualitative Analysis. J Psychoactive Drugs 2015;47(5):360-367. [doi: 10.1080/02791072.2015.1096434] [Medline: 26595229]
24. Adusumalli S, Lee H, Hoi Q, Koo S, Tan IB, Ng PC. Assessment of Web-Based Consumer Reviews as a Resource for Drug Performance. J Med Internet Res 2015 Aug 28;17(8):e211 [FREE Full text] [doi: 10.2196/jmir.4396] [Medline: 26319108]
25. Buysse DJ. Insomnia. JAMA 2013 Feb 20;309(7):706-716 [FREE Full text] [doi: 10.1001/jama.2013.193] [Medline: 23423416]
26. Bertisch SM, Herzig SJ, Winkelman JW, Buettner C. National use of prescription medications for insomnia: NHANES 1999-2010. Sleep 2014 Feb 01;37(2):343-349 [FREE Full text] [doi: 10.5665/sleep.3410] [Medline: 24497662]
27. Wilt TJ, MacDonald R, Brasure M, Olson CM, Carlyle M, Fuchs E, et al. Pharmacologic Treatment of Insomnia Disorder: An Evidence Report for a Clinical Practice Guideline by the American College of Physicians. Ann Intern Med 2016 Jul 19;165(2):103-112. [doi: 10.7326/M15-1781] [Medline: 27136278]
28. Kuriyama A, Honda M, Hayashino Y. Ramelteon for the treatment of insomnia in adults: a systematic review and meta-analysis. Sleep Med 2014 Apr;15(4):385-392. [doi: 10.1016/j.sleep.2013.11.788] [Medline: 24656909]
29. Coleman PJ, Gotter AL, Herring WJ, Winrow CJ, Renger JJ. The Discovery of Suvorexant, the First Orexin Receptor Drug for Insomnia. Annu Rev Pharmacol Toxicol 2017 Jan 06;57:509-533. [doi: 10.1146/annurev-pharmtox-010716-104837] [Medline: 27860547]
30. Drugs.com. Review drug. URL: https://www.drugs.com/members_comments_add/ [accessed 2019-01-15]
31. MedDRA. 2015. Introductory Guide MedDRA Version 18.0. URL: https://www.meddra.org/sites/default/files/guidance/file/intguide_18_0_english.pdf [accessed 2019-10-30]
32. FDA: U.S. Food and Drug Administration. 2018. FDA Adverse Event Reporting System (FAERS): Latest Quarterly Data Files. URL: https://www.fda.gov/drugs/guidancecomplianceregulatoryinformation/surveillance/adversedrugeffects/ucm082193.htm [accessed 2019-01-09]
33. Price J. What Can Big Data Offer the Pharmacovigilance of Orphan Drugs? Clin Ther 2016 Dec;38(12):2533-2545. [doi: 10.1016/j.clinthera.2016.11.009] [Medline: 27914633]
34. Gunja N. In the Zzz zone: the effects of Z-drugs on human performance and driving. J Med Toxicol 2013 Jun;9(2):163-171 [FREE Full text] [doi: 10.1007/s13181-013-0294-y] [Medline: 23456542]
35. FDA: U.S. Food and Drug Administration. 2019. FDA adds Boxed Warning for risk of serious injuries caused by sleepwalking with certain prescription insomnia medicines. URL: https://www.fda.gov/drugs/drug-safety-and-availability/fda-adds-boxed-warning-risk-serious-injuries-caused-sleepwalking-certain-prescription-insomnia [accessed 2019-08-11]

36. Sinha MS, Freifeld CC, Brownstein JS, Donneyong MM, Rausch P, Lappin BM, et al. Social Media Impact of the Food and Drug Administration's Drug Safety Communication Messaging About Zolpidem: Mixed-Methods Analysis. JMIR Public Health Surveill 2018 Jan 05;4(1):e1 [FREE Full text] [doi: 10.2196/publichealth.7823] [Medline: 29305342]
37. Misu T, Kortepeter CM, Muñoz MA, Wu E, Dal Pan GJ. An Evaluation of "Drug Ineffective" Postmarketing Reports in Drug Safety Surveillance. Drugs Real World Outcomes 2018 Jun;5(2):91-99 [FREE Full text] [doi: 10.1007/s40801-018-0131-3] [Medline: 29476420]
38. FDA.gov. Highlights of Prescribing Information. URL: https://www.accessdata.fda.gov/drugsatfda_docs/label/2014/204569s000lbledt.pdf [accessed 2019-01-09]
39. White RW, Harpaz R, Shah NH, DuMouchel W, Horvitz E. Toward enhanced pharmacovigilance using patient-generated data on the internet. Clin Pharmacol Ther 2014 Aug;96(2):239-246 [FREE Full text] [doi: 10.1038/clpt.2014.77] [Medline: 24713590]
40. Yom-Tov E, Gabrilovich E. Postmarket drug surveillance without trial costs: discovery of adverse drug reactions through large-scale analysis of web search queries. J Med Internet Res 2013;15(6):e124 [FREE Full text] [doi: 10.2196/jmir.2614] [Medline: 23778053]
41. Drugs.com. User Reviews for Olanzapine. URL: https://www.drugs.com/comments/olanzapine/ [accessed 2019-01-09]
42. Drugs.com. User Reviews for Methotrexate. URL: https://www.drugs.com/comments/methotrexate/ [accessed 2019-01-09]
43. Drugs.com. User Reviews for Clindamycin. URL: https://www.drugs.com/comments/clindamycin/ [accessed 2019-01-09]
44. Weldring T, Smith SMS. Patient-Reported Outcomes (PROs) and Patient-Reported Outcome Measures (PROMs). Health Serv Insights 2013;6:61-68 [FREE Full text] [doi: 10.4137/HSI.S11093] [Medline: 25114561]
45. Acquadro C, Berzon R, Dubois D, Leidy NK, Marquis P, Revicki D, et al. Incorporating the patient's perspective into drug development and communication: an ad hoc task force report of the Patient-Reported Outcomes (PRO) Harmonization Group meeting at the Food and Drug Administration, February 16, 2001. Value Health 2003;6(5):522-531 [FREE Full text] [doi: 10.1046/j.1524-4733.2003.65309.x] [Medline: 14627058]
46. Calvert M, Blazeby J, Altman DG, Revicki DA, Moher D, Brundage MD, CONSORT PRO Group. Reporting of patient-reported outcomes in randomized trials: the CONSORT PRO extension. JAMA 2013 Feb 27;309(8):814-822. [doi: 10.1001/jama.2013.879] [Medline: 23443445]
47. Basch E. The missing voice of patients in drug-safety reporting. N Engl J Med 2010 Mar 11;362(10):865-869 [FREE Full text] [doi: 10.1056/NEJMp0911494] [Medline: 20220181]
48. Buysse DJ, Yu L, Moul DE, Germain A, Stover A, Dodds NE, et al. Development and validation of patient-reported outcome measures for sleep disturbance and sleep-related impairments. Sleep 2010 Jun;33(6):781-792 [FREE Full text] [Medline: 20550019]
49. Roehrs T, Verster JC, Koshorek G, Withrow D, Roth T. How representative are insomnia clinical trials? Sleep Med 2018 Nov;51:118-123 [FREE Full text] [doi: 10.1016/j.sleep.2018.06.003] [Medline: 30144651]
50. Hughes S, Cohen D. Can online consumers contribute to drug knowledge? A mixed-methods comparison of consumer-generated and professionally controlled psychotropic medication information on the internet. J Med Internet Res 2011;13(3):e53 [FREE Full text] [doi: 10.2196/jmir.1716] [Medline: 21807607]
51. Kumar A. The Newly Available FAERS Public Dashboard: Implications for Health Care Professionals. Hosp Pharm 2019 Apr;54(2):75-77. [doi: 10.1177/0018578718795271] [Medline: 30923396]
52. Chew SW, Khoo CS. Comparison of drug information on consumer drug review sites versus authoritative health information websites. J Assn Inf Sci Tec 2015 Feb 13;67(2):333-349. [doi: 10.1002/asi.23390]
53. Carney CE, Buysse DJ, Ancoli-Israel S, Edinger JD, Krystal AD, Lichstein KL, et al. The consensus sleep diary: standardizing prospective sleep self-monitoring. Sleep 2012 Feb 01;35(2):287-302 [FREE Full text] [doi: 10.5665/sleep.1642] [Medline: 22294820]
54. Karapetiantz P, Bellet F, Audeh B, Lardon J, Leprovost D, Aboukhamis R, et al. Descriptions of Adverse Drug Reactions Are Less Informative in Forums Than in the French Pharmacovigilance Database but Provide More Unexpected Reactions. Front Pharmacol 2018;9:439. [doi: 10.3389/fphar.2018.00439] [Medline: 29765326]
55. Coloma PM, Becker B, Sturkenboom MCJM, van Mulligen EM, Kors JA. Evaluating Social Media Networks in Medicines Safety Surveillance: Two Case Studies. Drug Saf 2015 Oct;38(10):921-930 [FREE Full text] [doi: 10.1007/s40264-015-0333-5] [Medline: 26242616]

Abbreviations

FAERS: Food and Drug Administration Adverse Event Reporting System
LLT: low-level term
MedDRA: Medical Dictionary for Regulatory Activities
PT: preferred term

Edited by G Eysenbach; submitted 15.01.19; peer-reviewed by S Hughes, D Cohen, G Gonzalez, MS Aslam; comments to author 27.04.19; revised version received 22.08.19; accepted 26.09.19; published 08.11.19.

Please cite as:
Borchert JS, Wang B, Ramzanali M, Stein AB, Malaiyandi LM, Dineley KE
Adverse Events Due to Insomnia Drugs Reported in a Regulatory Database and Online Patient Reviews: Comparative Study
J Med Internet Res 2019;21(11):e13371
URL: http://www.jmir.org/2019/11/e13371/
doi:10.2196/13371
PMID:31702558

Original Paper

Patient-Reported Outcomes in Online Communications on Statins, Memory, and Cognition: Qualitative Analysis Using Online Communities

Farris Timimi[1], MD; Sara Ray[2], MA; Erik Jones[2], DMA; Lee Aase[1], BS; Kathleen Hoffman[2], PhD

[1]Department of Cardiovascular Disease, Mayo Clinic, Rochester, MN, United States

[2]Inspire, Arlington, VA, United States

Corresponding Author:
Farris Timimi, MD
Department of Cardiovascular Disease
Mayo Clinic
200 First Street, SW
Rochester, MN, 55905
United States
Phone: 1 507 284 1446
Email: timimi.farris@mayo.edu

Abstract

Background: In drug development clinical trials, there is a need for balance between restricting variables by setting eligibility criteria and representing the broader patient population that may use a product once it is approved. Similarly, although recent policy initiatives focusing on the inclusion of historically underrepresented groups are being implemented, barriers still remain. These limitations of clinical trials may mask potential product benefits and side effects. To bridge these gaps, online communication in health communities may serve as an additional population signal for drug side effects.

Objective: The aim of this study was to employ a nontraditional dataset to identify drug side-effect signals. The study was designed to apply both natural language processing (NLP) technology and hands-on linguistic analysis to a set of online posts from known statin users to (1) identify any underlying crossover between the use of statins and impairment of memory or cognition and (2) obtain patient lexicon in their descriptions of experiences with statin medications and memory changes.

Methods: Researchers utilized user-generated content on Inspire, looking at over 11 million posts across Inspire. Posts were written by patients and caregivers belonging to a variety of communities on Inspire. After identifying these posts, researchers used NLP and hands-on linguistic analysis to draw and expand upon correlations among statin use, memory, and cognition.

Results: NLP analysis of posts identified statistical correlations between statin users and the discussion of memory impairment, which were not observed in control groups. NLP found that, out of all members on Inspire, 3.1% had posted about memory or cognition. In a control group of those who had posted about TNF inhibitors, 6.2% had also posted about memory and cognition. In comparison, of all those who had posted about a statin medication, 22.6% (P<.001) also posted about memory and cognition. Furthermore, linguistic analysis of a sample of posts provided themes and context to these statistical findings. By looking at posts from statin users about memory, four key themes were found and described in detail in the data: memory loss, aphasia, cognitive impairment, and emotional change.

Conclusions: Correlations from this study point to a need for further research on the impact of statins on memory and cognition. Furthermore, when using nontraditional datasets, such as online communities, NLP and linguistic methodologies broaden the population for identifying side-effect signals. For side effects such as those on memory and cognition, where self-reporting may be unreliable, these methods can provide another avenue to inform patients, providers, and the Food and Drug Administration.

(J Med Internet Res 2019;21(11):e14809) doi:10.2196/14809

KEYWORDS

social media; hydroxymethylglutaryl-CoA reductase inhibitors; drug-related side effects and adverse reactions; memory loss; PROMs; pharmacovigilance; infodemiology; infoveillance; peer-support groups

Introduction

Background

Upon the implementation of the American College of Cardiology–American Heart Association's guidelines published in 2013, it was estimated that 1 billion people worldwide would be eligible to take statins to prevent cardiovascular diseases [1,2]. In fact, hydroxymethylglutaryl–coenzyme A reductase inhibitors, or statin medications, have become one of the most commonly prescribed classes of medications in the United States. Often positioned as safe and effective, their prevalence and relative acceptance do not mean that patients are without risk. Commonly cited statin-associated symptoms include muscular complaints (pain, cramps, and muscle weakness) [3-5], diabetes mellitus [6,7], and changes in memory or cognition [8-12].

Examples of research investigating the effects of statins on the central nervous system include case studies [8], observational studies [9], and randomized clinical trials [10-12] with meta-analyses reporting contradictory results [13,14]. Unfortunately, a discrepancy lies between the reported degree of memory changes and the actual observed frequency in large-scale clinical trials.

However, in 2012, on the basis of a review of spontaneous reports of amnesia, confusion, and concentration complaints, among others, from the Food and Drug Administration's (FDA) Adverse Event Reporting System, the FDA required a statin label change [15]:

> *Memory loss and confusion have been reported with statin use. These reported events were generally not serious and went away once the drug was no longer being taken.*

Many feel the evidence is inconclusive and the decision remains controversial.

In drug development clinical trials, there is a need for balance between restricting variables by setting eligibility criteria and representing the broader patient population that may use a product once it is approved. Similarly, although recent policy initiatives focusing on the inclusion of historically underrepresented groups are being implemented, barriers still remain. These limitations of clinical trials may mask potential product benefits and side effects. To bridge these gaps, online communication in health communities may serve as an additional population signal for drug side effects.

Using these novel data sources requires unique strategies. There is a wealth of patient experience data to be found in online patient forums. The data here are unstructured, allowing for naturally occurring themes and topics to be identified through organic patient and caregiver language.

Natural language processing (NLP) as well as hands-on linguistic analysis can be applied to online posts. In this study, posts were authored by patients and caregivers on Inspire, a company that creates and manages online support communities for more than 1 million patients and caregivers. Analysis of online patient and caregiver communications discussing statins utilized NLP methodology and technology to draw correlations among discussions of memory events by statin users and compared these events with discussions of memory events of patients using other classes of medication as well as discussions of memory events of all other patients on Inspire. The NLP system uses tokenization, lemmatization, stemming, edit distances, acronym dissection, and word and phrase boundaries to understand the content and meaning of the posts. These findings were then associated with Wikipedia entries and correlated with a dictionary of conditions and treatments from the National Institutes of Health (developed in partnership with Stanford University) to accurately extract the entities used in this study. In addition, manual curation of posts using linguistic analysis provided qualitative context to these statistical findings.

By applying these tools to posts created by members on Inspire who belonged to communities focused on heart health, the data could be used to look at statin medications and identify any underlying crossover between the use of statins and impairment of memory or cognition. Moreover, this combined strategy accessed the patient's voice, providing a detailed guide to how community members describe their experiences with statin medications and memory changes.

Objective

The objective of this study was to employ a nontraditional dataset to identify drug side-effect signals. Specifically, the study was designed to apply both NLP technology and hands-on linguistic analysis to a set of online posts from known statin users (1) to identify any underlying crossover between the use of statins and impairment of memory or cognition and (2) to obtain patient lexicon in their descriptions of experiences with statin medications and memory changes.

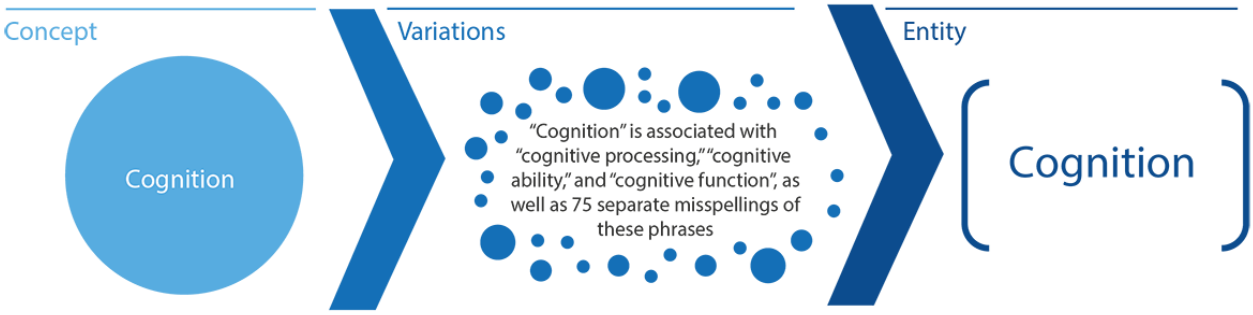
Methods

Researchers utilized user-generated content (UGC) on Inspire, looking at over 11 million posts across it. Posts were written by patients and caregivers belonging to a variety of communities on Inspire. After identifying these posts, researchers used NLP and hands-on linguistic analysis to draw correlations between statin use and memory and cognition.

Natural Language Processing and Statistical Comparison

In the corpus of over 11 million unique posts and over 440,000 different posters (patients and caregivers—*authors* of posts) on Inspire, the researchers used an NLP system that extracted relevant entities from each post. Entities are words or phrases that relate to each concept under consideration. For example, if the concept under consideration is cognition, several words or phrases related to that concept result in the creation of an entity (see Figure 1).

Figure 1. Creation of entities. The basis for entity extraction comes from Wikipedia and Wikidata, using technical tools created by the natural language processing company TextRazor.



Legend Figure 1. The basis for entity extraction comes from Wikipedia and Wikidata, using technical tools created by the NLP company TextRazor

The first stage of the analysis was to extract every post that contained an entity that related to memory loss or cognitive decline. The following are the entities that were used, each one of which has a distinct entry on Wikipedia: cognition, cognitive disorder, memory, recall, short-term memory, memory disorder, working memory, memory span, severe cognitive impairment, mild cognitive impairment, and cognitive deficit.

Note that each entity is associated with dozens of variations, that is, different phrasings, colloquialisms, and misspellings. For instance, *cognition* is associated with *cognitive processing*, *cognitive ability*, and *cognitive function*, as well as 75 separate misspellings of these phrases.

The researchers then did the same with statins, using the following entities: ulinastatin, cilastatin, niacin, simvastatin, fluvastatin, migrastatin, follistatin, mevastatin, oncostatin M, cystatin, cerivastatin, somatostatin, simvastatin, lovastatin, combretastatin A4 phosphate, myostatin, nystatin, atorvastatin, amlodipine, angiostatin, Crestor, Lipitor, and Vytorin.

Finally, the team identified medication discussions that could serve as the control group. The researchers wanted to choose a baseline of posts that indicated people were writing about a specific medication to reduce the experimental variation to a single change, specifically the medication being discussed. In this case, the decision was made to use Tumor Necrosis Factor TNF inhibitors.

TNF inhibitors were chosen for this statistical comparison for several reasons. First, TNF inhibitors are a commonly prescribed medication, used by a significant portion of Inspire members. Second, TNF inhibitors are prescribed for conditions that do not overlap with the conditions that statins treat; therefore, there would be as little overlap as possible between people who take both types of medications. Third, TNF inhibitors have shown no association with memory loss or cognitive decline. Fourth, generally the same age cohort uses TNF inhibitors and statins, minimizing age-related effects of cognitive decline between the 2 groups.

The entities used for TNF inhibitors were the following: Humira, golimumab, Enbrel, certolizumab pegol, and Remicade.

By finding authors who had written about one or more entities from each set and authors who had written posts that contained overlap among entities in multiple sets, the researchers were able to perform significant statistical analysis among the sets. It was irrelevant to the analysis as to what was mentioned first. An author could mention cognitive issues at one point in time and then later mention statins, or it could happen in the reverse order.

Hands-On Linguistic Analysis

To provide qualitative context to statistical findings, manual curation using linguistic methodology was performed on a subset of 246 UGC posts that mentioned both statin use and memory events. Posts were extracted and placed within an analyzable Excel file containing the following information: anonymized user identifier, date of post, title of the post, link to the post, and content of the post. Researchers developed a data-driven codebook with code label, full definition, and an example. Following the process described by Boyatzi, the team first reviewed and reduced the raw information; second, identified subsample themes; third, compared themes; fourth, created codes; and fifth, determined the reliability of the codes [16]. This comprised tags around top topics, challenges, and descriptors to identify patient themes around the type of memory impairment, experience of impairment, and the patient lexicon. Overall, 2 separate researchers applied the codebook to the data. Calculating reliability as the number of agreements divided by the total number of agreements plus disagreements resulted in intercoder reliability of 90%.

Ethics Statement

The study was reviewed internally by Mayo Clinic and found to be exempt from the review the of Institutional Review Board. All personally identifiable information was removed. All data were evaluated without the knowledge of the identity of those involved.

Results

NLP analysis of posts identified statistical correlations between statin users and discussion of memory impairment, which were not seen in control groups. Furthermore, linguistic analysis of a sample of posts provided themes and context to these statistical findings.

Natural Language Processing and Statistical Comparison

Through the analysis described above, the researchers found the following results for the number of people who had posted about one of these topics on Inspire (see Table 1).

To test the statistical significance of the observed high proportion of members that had a post that included the statins entity and that had a post that included memory, Fisher exact test (calculated in R, version 3.4.1) was applied. The calculated odds ratio was 9.703 (*P* value <.001), with a 95% CI of 9.066 to 10.378. This is outlined in a 2-way contingency table (Table 2) used to calculate the odds ratio and the associated significance measures.

Table 1. Inspire member numbers by post topic.

Inspire member numbers	Posts about anything	Posts about Tumor Necrosis Factor inhibitors	Posts about statin medication
Total members, n	440,835	14,323	5259
Subset of members who posted about memory, n (%)	13,878 (3.15)[a]	884 (6.17)[a]	1186 (22.55)[a]

[a]Percentage of overlap.

Table 2. Two-way contingency table calculating odds ratio and significance measures.

Entities	Memory entity	No memory entity	Sum
Statins entity	1186	4073	5259
No statin entity	12,692	422,884	435,576
Sum	13,878	426,957	440,835

Hands-On Linguistic Analysis

A total of 4 key themes related to the authors' memory were found in the data: memory loss, aphasia, cognitive impairment, and emotional change. Summaries of these themes, patient lexicon, and specific examples can be found in Table 3. Within these themes, descriptions of memory loss and confusion often overlap, and authors may label moments of cognitive impairment, confusion, or difficulties with function as memory loss. Regardless of the labels, authors were passionate about memory and cognitive impairment. Many used qualifiers of severity and speed, such as *catastrophic* or *instant*, when describing effects on their memory. Furthermore, a subset described their experience as *dementia* or *Alzheimer's*.

Memory difficulties were also attributed to aging, *pumphead* (postperfusion syndrome), surgery, or other medications. Authors who did attribute statin use to memory loss believed higher doses to be riskier. Authors also believed that their cognitive changes would resolve when statin use was discontinued. Nevertheless, during use, many memory events may have been written off by patients or may have gone unreported. Compounding the problem, authors reported having difficulty discussing their memory or cognitive issues with the doctors. Authors reported that many of their health care providers were dismissive of their claims, feeling that the benefits of statin medications far outweigh the risks.

Table 3. Patient lexicon and examples of key memory themes.

Key memory theme	Patient lexicon	Example
Memory loss: Though some patients have trouble with long-term memory, these patients are most passionate about and focused on the impact on short-term memory.	• Short-term/long-term memory loss • Reduced short-term memory • Short-term memory comes and goes • Memory problems • Memory issues • Memory impaired • Memory difficulties • Memory was shot • Decline in memory • Forgetful • Trouble remembering things	• Forgetting things that occurred, had been done, or were supposed to be done in the future • Forgetting things/facts authors are supposed to know
Aphasia: Authors describe difficulty communicating their thoughts. In particular, authors have difficulty remembering names of people that they recognize and know well.	• Verbal thinking • Loss of words • Word finding or unable to find words	• Difficulty recalling words, particularly names • Difficulty forming sentences
Cognitive impairment: Authors describe a loss of ability to think, reason, or understand. Authors are also concerned with the loss of ability to function. Authors are particularly concerned about attention span and forgetting how to do basic tasks.	• Confusion/mental confusion • Alzheimer's/instant Alzheimer's • Dementia • Brain fog/foggy • Cognitive loss • Neurological issues • Fuzzy thinking • Slow • Cognitive problems • Senile • Mental focus • Cognitive impairment	• Loss of focus or attention span • Inability to type or write • Difficulty counting money or shopping • Tasks take longer to accomplish or forgetting how to do a task • Getting into the wrong car or being unable to identify one's car • Wandering around in the night and not knowing why • Inability to recognize people • Cannot recognize known people • Difficulty explaining things to others, particularly doctors
Emotional change: Authors describe their emotions as changed, struggling with feeling depressed or lack of emotion. Authors also describe becoming angry easily or being moody or disinterested. A small subset feels increased anxiety.	• Depression/sad • Quick temper • Moody/moodiness • Anxious • Tired • Emotional lability	• Loss of desire or trouble feeling happy • Overreacting to situations • Becoming angry quickly • Worry • Feelings of ambivalence

Discussion

Principal Findings

This study uses online communities of self-identified patients and caregivers to evaluate the signals of self-reported memory impairment in statin users. A cohort of statin users was compared with a cohort of patients using TNF inhibitors and also to the patients on the site overall. Though discussions of memory impairment occurred 3.1% and 6.2% of the time in the overall population and the TNF inhibitor users, respectively, overlap between the statin user conversation and memory impairment posts was 22.6%, indicating a much higher and significantly different correlation between those on statin medications and the discussion of memory issues.

Furthermore, linguistic analysis was performed on a set of posts from statin users that discussed memory impairment to identify key themes. These patients and caregivers identified difficulty with memory loss, aphasia, cognitive function, and emotional change. Patients and caregivers indicated the speed and severity of these changes, likening their experience to *instant Alzheimer's*.

Although there are clinical trials that have not found associations of cognitive impairment with statin use, those trials were specifically conducted to evaluate cardiovascular outcomes, not cognitive outcomes [17-19]. During the postmarketing phase, a double-blind study comparing statins with placebo found small but significant differences in neuropsychological tests on attention and psychomotor speed [11]. Using the Naranjo adverse drug reaction (ADR) probability scale in a survey of 171 patients on statins, 75% of the participants were found to have experienced cognitive ADRs [20]. Another analysis involved examination of the FDA's Adverse Event Reporting System. It revealed a significantly higher proportion of adverse reports for statins as compared with control medications [21].

This study's findings support the importance of using novel and nontraditional data sources as additional population signals for side effects. The nature of memory issues and cognitive impairment can lead to underreporting of these issues to doctors. Furthermore, in the online posts, patients report doctors discounting, ignoring, or not taking their concerns over memory issues seriously. Improved patient care and outcomes are seriously impaired when their concerns are discounted. Underreporting of this side effect to the FDA is inevitable when patient complaints and concerns are not addressed. This

underscores the imperative of bringing the patient experience to light in nontraditional ways.

Drug development clinical trials should balance eligibility criteria, which allow for a defined population to be studied, against the limitations these criteria create. Specifically, eligibility criteria exclude and narrow, reducing the representativeness of the data. This subset of population may not represent the broader patient population that may use a product once it is approved. Although there have been recent policy initiatives that focus on the inclusion of historically underrepresented groups, barriers still remain that may limit or mask the potential benefits and side effects, respectively. Data provided from online communities may allow for an expansion of cohorts for study, potentially accessing patients who may not have traditionally been enrolled in other forms of research. Analyzing data from these patients and using mixed quantitative and qualitative methodologies allows corroboration of findings and provides a new avenue of side effect identification and investigation.

Limitations

This study had some limitations. First, clinical diagnosis, medical history, and current treatment of individual authors are self-reported, cannot be confirmed, and may be missing some information. Second, demographic data of the authors are unknown, making it unclear how the data reflect the general population. In addition, there may be an element of detection bias that may be relevant, given the intrinsic virality of online communication. Moreover, qualitative analysis may also reflect a limitation of outcome misclassification predicated upon word selection on the part of participants. Search terms were selected based on free-text review to limit the impact of alternative term selection. Despite this, missed cases may be a problem if community participants chose not to discuss or share memory challenges owing to perceived associated stigma.

One strength of qualitative research, detailed information about the human experience, makes it a compelling tool when applied to health [22]. The large datasets of online communication on social media sites, such as Inspire, can challenge the skills of individual researchers [23]. Maintenance of rigor in analysis may be impacted. Using NLP can effectively deal with this limitation.

However, in a methodological research study comparing NLP-only analysis, qualitative text–only analysis, and combined NLP–qualitative text analysis, researchers concluded that NLP-only analysis was an effective tool to identify and quantify major themes but lacked the ability to capture contextual nuances necessary for clarity and understanding. Combining the 2 methodologies provided the most comprehensive and highest quality results [24].

Comparison With Prior Work

Stanford University Medical School in collaboration with Inspire conducted research on over 8 million posts on Inspire utilizing NLP to look for associations between chemotherapeutic agents and ADRs. The research specifically extracted mentions of common and rare cutaneous ADRs (eg, rashes, blisters, and psoriasis flares) from posts related to (1) the epidermal growth factor receptor inhibitor, erlotinib, and (2) the immune checkpoint programmed cell death–1 inhibitors, nivolumab and pembrolizumab. The team discovered that some patients receiving the chemotherapy drug erlotinib (Tarceva) reported hypohidrosis—the inability to sweat—a condition that can lead to heat exhaustion, heat stroke, or even death. This ADR had never been reported in the medical literature, but Inspire members had been discussing it for over 11 years. The team also found that Inspire members discussed, among themselves, ADRs for other checkpoint inhibitors much earlier than reported in the medical literature—an average of 7 months before any of these side effects had been reported [25]. This research demonstrated the untapped resources and information to be uncovered in online health community postings.

Conclusions

Estimates of cessation of statin therapy within the first year are as high as 50%. This high degree of discontinuation is disconcerting as statins have been shown to reduce cardiovascular disease risk, accruing with each year of use and benefits persisting over the long term [26]. Side effects play a role in medication cessation. Although the benefits of statin medications may outweigh the risks, it is still worth identifying adverse reactions to better and more fully inform patient and doctor decisions.

Using the deidentified communication occurring online can add to the knowledge base about medication usage and adverse reactions. Large online populations, comprising large cohorts of patients who may not traditionally participate in research, can serve as additional population signals for side effects.

As with the prior work with erlotinib, this type of research can be used to inform the FDA of the side effects or adverse reactions that are presently unreported for a number of reasons. Using the combination of NLP analysis with qualitative analysis broadens and deepens learnings, tests hypotheses, and uncovers insights into the patient experience. Regarding statin use, patients may not connect their medications to cognitive changes, may not feel comfortable informing their physicians, or may be unable to articulate these changes in the clinical setting. If patients do tell the doctors, doctors may not relay this information to the FDA or feel that benefits outweigh the risks.

Further research is needed to truly identify the extent of cognitive change that is occurring with the use of statins.

Conflicts of Interest

None declared.

References

1. Goff DC, Lloyd-Jones DM, Bennett G, Coady S, D'Agostino RB, Gibbons R, American College of Cardiology/American Heart Association Task Force on Practice Guidelines. 2013 ACC/AHA guideline on the assessment of cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. Circulation 2014 Jun 24;129(25 Suppl 2):S49-S73. [doi: 10.1161/01.cir.0000437741.48606.98] [Medline: 24222018]
2. Ioannidis J. More than a billion people taking statins?: Potential implications of the new cardiovascular guidelines. J Am Med Assoc 2014 Feb 5;311(5):463-464. [doi: 10.1001/jama.2013.284657] [Medline: 24296612]
3. Pasternak RC, Smith SC, Bairey-Merz CN, Grundy SM, Cleeman JI, Lenfant C, American College of Cardiology, American Heart Association, National Heart, LungBlood Institute. ACC/AHA/NHLBI clinical advisory on the use and safety of statins. J Am Coll Cardiol 2002 Aug 7;40(3):567-572 [FREE Full text] [doi: 10.1016/s0735-1097(02)02030-2] [Medline: 12142128]
4. Thompson PD, Panza G, Zaleski A, Taylor B. Statin-associated side effects. J Am Coll Cardiol 2016 May 24;67(20):2395-2410 [FREE Full text] [doi: 10.1016/j.jacc.2016.02.071] [Medline: 27199064]
5. Bruckert E, Hayem G, Dejager S, Yau C, Bégaud B. Mild to moderate muscular symptoms with high-dosage statin therapy in hyperlipidemic patients--the PRIMO study. Cardiovasc Drugs Ther 2005 Dec;19(6):403-414. [doi: 10.1007/s10557-005-5686-z] [Medline: 16453090]
6. Ridker PM, Danielson E, Fonseca FA, Genest J, Gotto AM, Kastelein JJ, JUPITER Study Group. Rosuvastatin to prevent vascular events in men and women with elevated C-reactive protein. N Engl J Med 2008 Nov 20;359(21):2195-2207. [doi: 10.1056/NEJMoa0807646] [Medline: 18997196]
7. Freeman DJ, Norrie J, Sattar N, Neely RD, Cobbe SM, Ford I, et al. Pravastatin and the development of diabetes mellitus: evidence for a protective treatment effect in the West of Scotland Coronary Prevention Study. Circulation 2001 Jan 23;103(3):357-362. [doi: 10.1161/01.cir.103.3.357] [Medline: 11157685]
8. Wagstaff LR, Mitton MW, Arvik BM, Doraiswamy PM. Statin-associated memory loss: analysis of 60 case reports and review of the literature. Pharmacotherapy 2003 Jul;23(7):871-880. [doi: 10.1592/phco.23.7.871.32720] [Medline: 12885101]
9. Yaffe K, Barrett-Connor E, Lin F, Grady D. Serum lipoprotein levels, statin use, and cognitive function in older women. Arch Neurol 2002 Mar;59(3):378-384. [doi: 10.1001/archneur.59.3.378] [Medline: 11890840]
10. Golomb BA, Dimsdale JE, Koslik HJ, Evans MA, Lu X, Rossi S, et al. Statin effects on aggression: results from the UCSD statin study, a randomized control trial. PLoS One 2015;10(7):e0124451 [FREE Full text] [doi: 10.1371/journal.pone.0124451] [Medline: 26132393]
11. Muldoon MF, Barger SD, Ryan CM, Flory JD, Lehoczky JP, Matthews KA, et al. Effects of lovastatin on cognitive function and psychological well-being. Am J Med 2000 May;108(7):538-546. [doi: 10.1016/s0002-9343(00)00353-3] [Medline: 10806282]
12. Muldoon MF, Ryan CM, Sereika SM, Flory JD, Manuck SB. Randomized trial of the effects of simvastatin on cognitive functioning in hypercholesterolemic adults. Am J Med 2004 Dec 1;117(11):823-829. [doi: 10.1016/j.amjmed.2004.07.041] [Medline: 15589485]
13. McGuinness B, O'Hare J, Craig D, Bullock R, Malouf E, Passmore P. Statins for the treatment of dementia. Cochrane Database Syst Rev 2010 Aug 04(8):CD007514. [doi: 10.1002/14651858.CD007514.pub2] [Medline: 20687089]
14. Ott BR, Daiello LA, Dahabreh IJ, Springate BA, Bixby K, Murali M, et al. Do statins impair cognition? A systematic review and meta-analysis of randomized controlled trials. J Gen Intern Med 2015 Mar;30(3):348-358 [FREE Full text] [doi: 10.1007/s11606-014-3115-3] [Medline: 25575908]
15. The Food and Drug Administration. 2016. FDA Drug Safety Communication: Important Safety Label Changes to Cholesterol-lowering Statin Drugs URL: https://www.fda.gov/Drugs/DrugSafety/ucm293101.htm [accessed 2019-07-01]
16. Boyatzis R. Transforming Qualitative Information: Thematic Analysis and Code Development. Thousand Oaks: Sage; 1998.
17. Mora S, Ridker PM. Justification for the use of Statins in primary prevention: an Intervention Trial Evaluating Rosuvastatin (JUPITER)--can C-reactive protein be used to target statin therapy in primary prevention? Am J Cardiol 2006 Jan 16;97(2A):33A-41A. [doi: 10.1016/j.amjcard.2005.11.014] [Medline: 16442935]
18. Heart Protection Study Collaborative Group. MRC/BHF Heart Protection Study of cholesterol lowering with simvastatin in 20,536 high-risk individuals: a randomised placebo-controlled trial. Lancet 2002 Jul 6;360(9326):7-22. [doi: 10.1016/S0140-6736(02)09327-3] [Medline: 12114036]
19. Shepherd J, Blauw GJ, Murphy MB, Bollen EL, Buckley BM, Cobbe SM, PROSPER study group. PROspective Study of Pravastatin in the Elderly at Risk. Pravastatin in elderly individuals at risk of vascular disease (PROSPER): a randomised controlled trial. Lancet 2002 Nov 23;360(9346):1623-1630. [doi: 10.1016/s0140-6736(02)11600-x] [Medline: 12457784]
20. Evans MA, Golomb BA. Statin-associated adverse cognitive effects: survey results from 171 patients. Pharmacotherapy 2009 Jul;29(7):800-811. [doi: 10.1592/phco.29.7.800] [Medline: 19558254]
21. Sahebzamani FM, Munro CL, Marroquin OC, Diamond DM, Keller E, Kip KE. Examination of the FDA warning for statins and cognitive dysfunction. J Pharmacovigil 2014;2:4. [doi: 10.4172/2329-6887.1000141]
22. Anderson C. Presenting and evaluating qualitative research. Am J Pharm Educ 2010 Oct 11;74(8):141 [FREE Full text] [doi: 10.5688/aj7408141] [Medline: 21179252]

23. Bradley EH, Curry LA, Devers KJ. Qualitative data analysis for health services research: developing taxonomy, themes, and theory. Health Serv Res 2007 Aug;42(4):1758-1772 [FREE Full text] [doi: 10.1111/j.1475-6773.2006.00684.x] [Medline: 17286625]
24. Guetterman TC, Chang T, DeJonckheere M, Basu T, Scruggs E, Vydiswaran VV. Augmenting qualitative text analysis with natural language processing: methodological study. J Med Internet Res 2018 Jun 29;20(6):e231 [FREE Full text] [doi: 10.2196/jmir.9702] [Medline: 29959110]
25. Ransohoff JD, Nikfarjam A, Jones E, Loew B, Kwong BY, Sarin KY, et al. Detecting chemotherapeutic skin adverse reactions in social health networks using deep learning. JAMA Oncol 2018 Apr 1;4(4):581-583 [FREE Full text] [doi: 10.1001/jamaoncol.2017.5688] [Medline: 29494731]
26. Collins R, Reith C, Emberson J, Armitage J, Baigent C, Blackwell L, et al. Interpretation of the evidence for the efficacy and safety of statin therapy. Lancet 2016 Nov 19;388(10059):2532-2561. [doi: 10.1016/S0140-6736(16)31357-5] [Medline: 27616593]

Abbreviations

ADR: adverse drug reaction
FDA: Food and Drug Administration
NLP: natural language processing
TNF: Tumor Necrosis Factor
UGC: user-generated content

Edited by G Eysenbach; submitted 24.05.19; peer-reviewed by R Booth, J Brixey; comments to author 18.07.19; revised version received 30.08.19; accepted 24.09.19; published 28.11.19.

Please cite as:
Timimi F, Ray S, Jones E, Aase L, Hoffman K
Patient-Reported Outcomes in Online Communications on Statins, Memory, and Cognition: Qualitative Analysis Using Online Communities
J Med Internet Res 2019;21(11):e14809
URL: http://www.jmir.org/2019/11/e14809/
doi:10.2196/14809
PMID:31778117

Original Paper

Artificial Intelligence Technologies for Coping with Alarm Fatigue in Hospital Environments Because of Sensory Overload: Algorithm Development and Validation

Chrystinne Oliveira Fernandes[1*], MSc; Simon Miles[2], PhD; Carlos José Pereira De Lucena[1*], PhD; Donald Cowan[3*], PhD

[1]Department of Informatics, Pontifical Catholic University of Rio de Janeiro, Rio de Janeiro, Brazil

[2]Kings College London, London, United Kingdom

[3]University of Waterloo, Waterloo, ON, Canada

[*]these authors contributed equally

Corresponding Author:
Chrystinne Oliveira Fernandes, MSc
Department of Informatics
Pontifical Catholic University of Rio de Janeiro
Rio Datacenter, 4th Fl
225 Marquês de São Vicente St
Rio de Janeiro, 22451-900
Brazil
Phone: 55 21 3527 1510
Fax: 55 3527 1530
Email: chrystinne@gmail.com

Abstract

Background: Informed estimates claim that 80% to 99% of alarms set off in hospital units are false or clinically insignificant, representing a cacophony of sounds that do not present a real danger to patients. These false alarms can lead to an alert overload that causes a health care provider to miss important events that could be harmful or even life-threatening. As health care units become more dependent on monitoring devices for patient care purposes, the alarm fatigue issue has to be addressed as a major concern for the health care team as well as to enhance patient safety.

Objective: The main goal of this paper was to propose a feasible solution for the alarm fatigue problem by using an automatic reasoning mechanism to decide how to notify members of the health care team. The aim was to reduce the number of notifications sent by determining whether or not to group a set of alarms that occur over a short period of time to deliver them together, without compromising patient safety.

Methods: This paper describes: (1) a model for supporting reasoning algorithms that decide how to notify caregivers to avoid alarm fatigue; (2) an architecture for health systems that support patient monitoring and notification capabilities; and (3) a reasoning algorithm that specifies how to notify caregivers by deciding whether to aggregate a group of alarms to avoid alarm fatigue.

Results: Experiments were used to demonstrate that providing a reasoning system can reduce the notifications received by the caregivers by up to 99.3% (582/586) of the total alarms generated. Our experiments were evaluated through the use of a dataset comprising patient monitoring data and vital signs recorded during 32 surgical cases where patients underwent anesthesia at the Royal Adelaide Hospital. We present the results of our algorithm by using graphs we generated using the R language, where we show whether the algorithm decided to deliver an alarm immediately or after a delay.

Conclusions: The experimental results strongly suggest that this reasoning algorithm is a useful strategy for avoiding alarm fatigue. Although we evaluated our algorithm in an experimental environment, we tried to reproduce the context of a clinical environment by using real-world patient data. Our future work is to reproduce the evaluation study based on more realistic clinical conditions by increasing the number of patients, monitoring parameters, and types of alarm.

(J Med Internet Res 2019;21(11):e15406) doi:10.2196/15406

KEYWORDS

alert fatigue health personnel; health information systems; patient monitoring; alert systems; artificial intelligence

Introduction

Alarm Fatigue

Information Technology (IT) has already provided significant benefits to the health care sector, but there are still many areas where the application of IT could offer further critical improvements. For example, alarm fatigue, which has recently been receiving attention from industry, the health care sector, and the academic community, is a worldwide hospital problem.

Alarm fatigue involves a lack of response because of an excessive number of noncritical alarms being received by health care personnel, resulting in sensory overload and desensitization [1-4]. To illustrate the severity of this problem that has been treated as a major patient safety concern, scientific studies have reported that there was an average of 700 physiologic monitor alarms per patient per day [1]. Such a number indicates a severe sensory overload for the health care staff, with serious consequences for the well-being of the patients when an alarm might be ignored.

In this paper, we present a new approach to coping with the alarm fatigue problem, its most common causes, adverse consequences, and strategies as compared with other solutions published in the literature [5-9]. Our proposed solution for addressing this issue uses an artificial intelligence (AI) approach based on an automatic reasoning system that decides how to notify caregivers about anomalies detected by a patient monitoring system where a large volume of alarms could lead to alarm fatigue. In other words, we are using IT to reduce the number of notifications received by health care staff, so they can be focused on the activities that truly need attention. Our experiments were configured to alert nurses and were evaluated through the use of a dataset comprising a wide range of patient monitoring data and vital signs that were recorded during 32 surgical cases where patients underwent anesthesia at the Royal Adelaide Hospital [10].

In this work, we aim at addressing 2 main research questions: (1) How can an automatic reasoning system determine how to notify caregivers about anomalies detected by a patient monitoring system where a large volume of alarm leads to alarm fatigue? (2) How to reason about avoiding alarm fatigue?

Our main goal with this case study is to find out whether to group a set of alarms that happens within a short period of time to deliver them together without compromising patient safety. Our specific goal is to avoid that alarms of the same type for the same patient can be alerted more than once within a short period by using a notification delay strategy.

Theoretical Background

Related Work

A critical concern in hospitals that use monitoring devices to track patients' health is alarm fatigue. Tens of thousands of alerts may go off throughout a hospital each day, and yet some 80% to 99% of these audible or visual alerts are false or nuisance alarms, indicating conditions that do not require clinical intervention [1-4]. Alarm fatigue represents a substantial issue that can bring undesired consequences to health care environments. For instance, the desensitization of a health care team to alerts can lead to longer response times for handling anomalies as well as possibly missing life-threatening events. These examples illustrate the fact that sensory overload is very likely to produce an unsafe environment for patients.

According to Sowan et al [6], the key issues causing alarm fatigue and reducing trust in alarm systems are as follows: the high incidence of nuisance alarms, the confusion in locating the device sending out the alarm, unit layouts that hinder alarm response, the inadequacy of alarm systems to alert nurses of changes in patients' conditions, and the complexity of new monitoring systems, among others. The most important issues interfering with alarm recognition and alarm response ranked by the nurses in [6] were as follows: (1) frequent false alarms, (2) difficulty in understanding alarm priorities, and (3) noise competition from nonclinical devices.

Caring for patients and managing alarms simultaneously is a very complex and demanding task, especially when health providers are caring for multiple patients at the same time and have been exposed to a high number of alarms generated by physiological monitors. In addition to dealing with frequent alarms, health care providers also perform other activities, such as medication administration, patient assessments, and note updates. Over time, they become fatigued and errors may occur because of decreased attentiveness [5].

Considering the aforementioned scenario, a commonly recommended solution to mitigate alarm fatigue is to adjust alarm parameters on monitors to suit each patient's conditions rather than using default settings [5]. The works of Shanmugham et al [5] and Sowan et al [6] are examples of studies that assess the effect of modifying the default alarm settings provided by the device manufacturers. According to their findings, the nurses' perceived workload was lower when the clinical alarm threshold limits were modified according to patients' clinical conditions. They also concluded that the modification of alarm settings affects the number of alarms accurately addressed, care providers' experience, and overall satisfaction.

Another strategy suggested to reduce the number of false alarms and alarm fatigue is educating staff regarding alarm management [6]. Sowan et al showed that their changes in default alarm settings significantly reduced 24% of the total number of the target alarms after their interventions, which included the following: (1) re-education of intensive care unit (ICU) bedside nurses on the appropriate use of the monitors, and (2) changing default settings of some parameters on the cardiac monitors, including the addition of an alarm delay by increasing the period between the alarm detection and its triggering, among others.

However, despite the achievement of a significant reduction in the alarm rate, they deem that the changing of default settings

and better education regarding cardiac monitors are insufficient to improve alarm system safety.

Scientific studies show that the quality of medical device alarms is unsatisfactory, and it affects quality of care and patient safety. One root cause is the poor quality of alarm-generating algorithms. Therefore, from a clinical perspective, major improvements in alarm algorithms are urgently needed [8].

To pursue this goal, different methods have been proposed and investigated for use in the alarm systems of medical devices, mostly from the fields of statistics and AI. Imhoff et al gave a brief overview of different methods, including statistical approaches (eg, improved data preprocessing, robust signal extraction, segmentation, median filter, statistical process control, and time series analysis for pattern detection, among others) and AI methods, such as knowledge-based approaches, knowledge discovery based on machine learning, neural networks, random forests, fuzzy logic, and Bayesian networks [8].

Regarding the methodological approaches to alarm management, Imhoff et al present the 4 areas in which alarms can be improved: (1) signal acquisition, that is, the interface between patient and medical devices; (2) alarm generation, that is, the algorithms that determine an alarm situation; (3) alarm validation, that is, determining whether the alarm is actually valid; (4) integration of multiple alarms, for example, from different devices, into 1 or few alarms [8].

Successful quality improvement approaches included alteration in default monitor presets, daily electrode change, alarm customization, alarm management education, change in policy, histogram-based pulse oximetry (SpO_2), alarm tailoring, improved displays to aid in nurse-patient assignment, and the use of notification delays [10]. Notification delays are performed with a middleware situated between the alarming medical device and the clinicians' receiver equipment such as a mobile phone. Several studies found that introducing alarm delays prior to the notification process could drop "false alarms" 25–67% [10]. Regarding the reduction of the total alarms, considering the effects of these interventions, alarm quantities decreased between 18.5% and as much as 89%, according to Winters et al [9].

The major contribution of our work described in this paper is mainly related to the integration/grouping of multiple alarms, where we present the application of a new alarm algorithm to reduce alarm fatigue. We evaluated our algorithm to reduce the total number of alarms through the use of real patient data. Our approach uses a notification delay approach to decide whether to deliver a unique notification to caregivers instead of several alarms for the same alarm situation. By using our system, we reduced the notifications received by the caregivers by up to 99.3% (582/586) of the total alarms generated.

A Step Before Reasoning: The Anomaly Detection and Alarm-Triggering Processes

Before presenting our reasoning mechanism, we outline important concepts of the monitoring process developed in our previous work related to coping with remote patient monitoring [11,12]. In this section, we illustrate a more formal model for the anomaly detection and the alarm-triggering processes that are used in our system.

The default functioning of our notification system is to notify a group of caregivers about anomalies detected in a patient's vital signs. The anomaly detection process works through continuous monitoring of each patient's vital signs. To verify if an anomaly occurs, the readings are evaluated against anomaly thresholds configured for each patient. If a reading for a patient is more than a maximum or less than a minimum threshold value, then the reading is considered to be anomalous and the system triggers an alarm that is sent to the health care team. The anomaly detection process and its related concepts such as anomalies, alarms, and notifications are defined in the next subsections.

Defining Thresholds and Anomalous Values

Anomaly thresholds for the sensors must be configured before starting to monitor a patient. A threshold is a minimum and maximum limit for a reading of a sensor S for a patient P, and an anomaly is a value either below or above those limits. An anomaly or anomalous value (AV) $v \in AV(S,P)$ triggers an alarm that is sent to the health care team. The threshold value for sensor S connected to a patient P is designated threshold (S, P) and the minimum and maximum values are $v\_\min(S, P)$ and $v\_\max(S, P)$, respectively. We formally defined anomalies using set theory as shown later.

Let $AV(S,P)$ be the set of values that represent patient P's AV for the sensor S. Let us also consider that these values from S belong to the set of real numbers. The $AV(S,P)$ set is formally defined as shown in Equation (1):

$$AV_{S,P}=\{v \mid v \in R,\ v\_\min_{S,P} > v > v\_\max_{S,P}\} \quad (1)$$

Where:

1. The inequalities v<v_minS,P and v>v_maxS,P comprise the thresholds for sensor S and patient P.
2. v_minS,P∈ R, which represents the minimum limit, that is, the value below which a sensor reading v is considered an AV.
3. v_maxS,P∈ R, representing the maximum limit, that is, the value above which v is considered an anomaly.

We can define an anomaly detected by sensor as the function $An(v)=b$ that maps real numbers into Booleans (f: R→ Boolean) where v∈ R and is the value that represents a sensor reading, and b={true, false} as shown below.

An(v)=true, if v∈ AVS,P; false, otherwise. (2)

Defining Alarm, Anomaly Detection, and Notification Events

In our system, we define the concepts of anomaly detection, alarm triggering, and notification in terms of events, which are represented as α, β, and μ, respectively.

The occurrence of an event α="anomaly detected" means that the function $An(v)$ assumes the value "true" at a given time defined as ANOMALY_DETECTION_TIME (T_α). The event β="alarm triggering," in its turn, is defined as the action of

triggering an alarm to indicate that an anomaly has been detected. The time when an event β occurs is referred as ALARM_TRIGGERING_TIME (T_β). The third event we define in this section is μ="notification." μ is the action of sending a notification to a set of caregivers to inform them that an alarm has been triggered. The time when an event μ occurs is referred to as NOTIFICATION_TIME (T_μ).

Associated with the occurrence of these events, we have the delays ALARM_TRIGGERING_DELAY (D_β) and NOTIFICATION_DELAY (D_μ), where D_β represents the delay between anomaly detection and its indication through an alarm triggering and D_μ is the delay between an alarm triggering and its notification to the caregivers. We show in Equations (3) and (4) how the delays D_β and D_μ are calculated according to the time at which the events α, β, and μ occur.

$$D\beta = T\beta - T\alpha \quad (3)$$

$$D\mu = T\mu - T\beta \quad (4)$$

We can summarize the abovementioned explanation in a more formal way through the event-trigger rules presented in Equations (5) and (6):

$$\phi 1: \alpha \rightarrow \beta \quad (5)$$

$$\phi 2: \beta \rightarrow \mu \quad (6)$$

where α, β, and μ are the events; the symbol "→" represents the action triggers; ϕ1 indicates that, when the event α occurs, the event β is automatically triggered after the delay D_β; and ϕ2 indicates that event is automatically triggered D_μ time after β occurs.

The parameterization of the events α, β, and μ is defined as follows.

$$A = <type, T\alpha> \quad (7)$$

$$B = <type, \alpha, T\beta> \quad (8)$$

$$M = <type, \beta, T\mu> \quad (9)$$

where the parameter α for β event represents the event α; and the parameter β for μ event represents the event "alarm triggering" β.

Modeling Anomaly Detection, Alarm-Triggering, and Notification

To illustrate the anomaly detection, alarm-triggering, and notification processes, we present a state-transition diagram in Figure 1. This figure presents a visual representation of the following: (1) the possible states of the anomaly detection, alarm-triggering, and notification processes; (2) the events such as inputs that may result in transitions between states; and (3) the transitions between states. We also show the conditions an event requires to trigger a transition.

To formalize the concept of an anomaly, we present, through the state-transition machine in Figure 2, the possible states for an anomaly. Figure 2 presents the current anomaly detection process, showing the 3 possible states of an anomaly: *no anomaly*, *anomaly alerted*, and *anomaly notified*. The interconnecting arrows represent the transitions between states, and the labels on the arrows represent the events that make the transitions occur.

Figure 1. The state-transition machine showing the states involved in the anomaly detection, alarm-triggering, and notification processes.

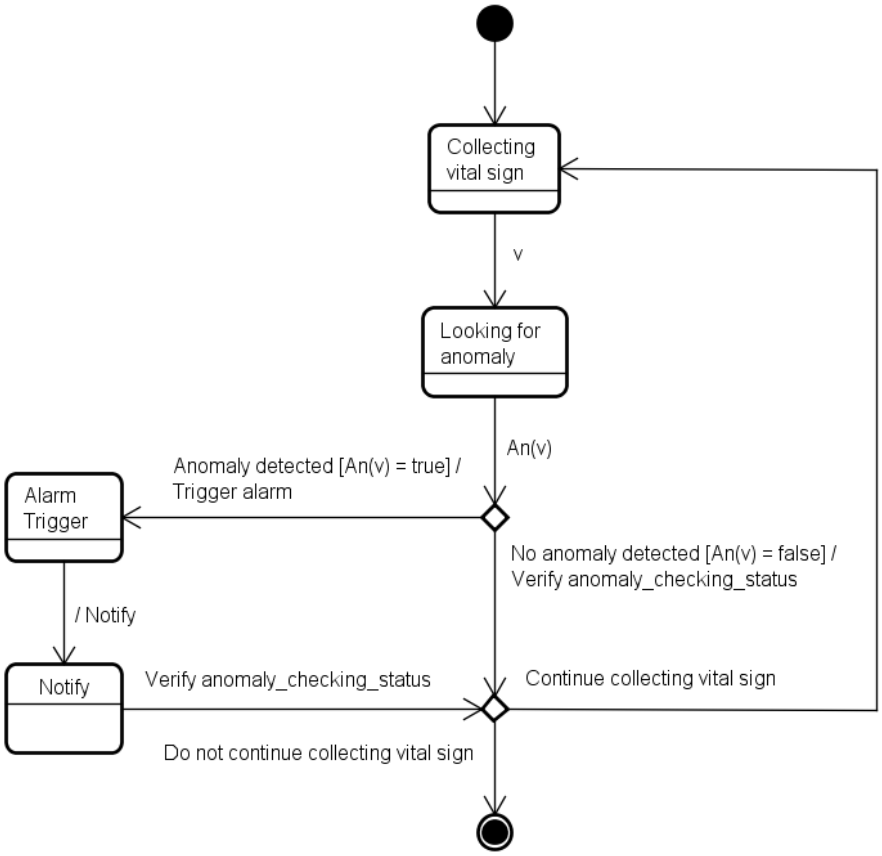
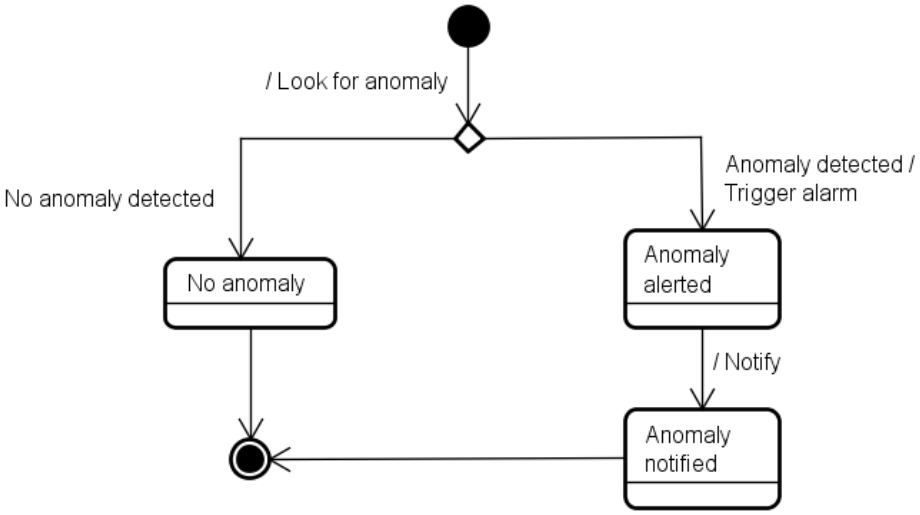


Figure 2. The state-transition machine showing the possible states for an anomaly.



Now that the basic concepts anomaly, alarm, anomaly detection, and notification needed for the reasoning process have been defined, in the next sections we present our reasoning model, system architecture, and algorithms.

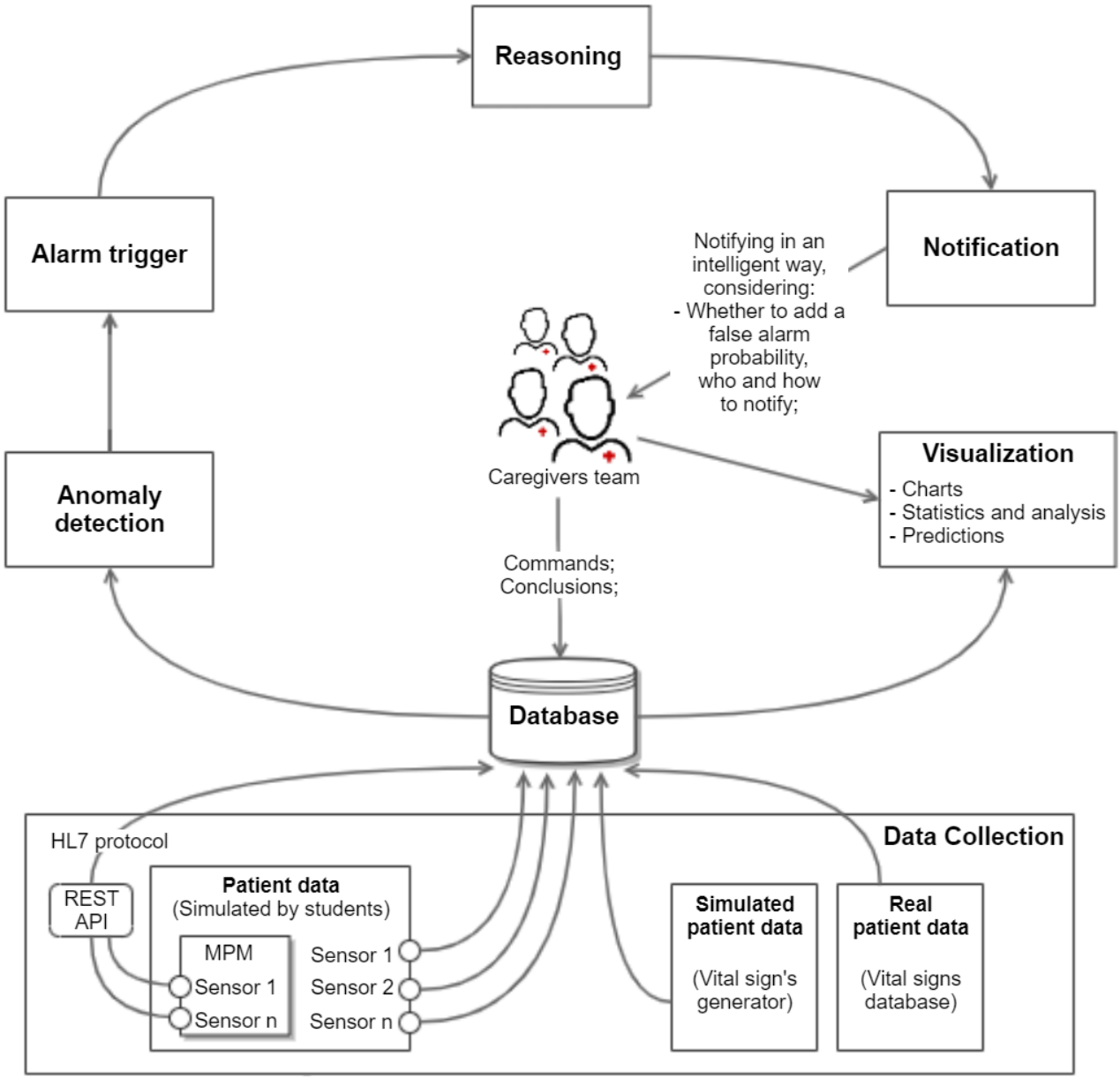
Adding Reasoning to the System

In this section, we provide a brief description of how we apply a reasoning engine to the alarms generated by the monitoring devices being used to track a patient's health status to minimize alarm fatigue. The software system contains a component that reads the vital signs (the reader) accompanied by a reasoning engine that decides how to notify the health care team. The reader can be set to ignore all the nonanomalous vital signs to focus only on the AV that can require attention from the caregivers' team. An anomalous reading is then passed to the reasoning engine that decides how to handle the reading. For example, the reading could be used to cause an alarm to be triggered immediately because the patient's situation is deemed critical; or readings could be accumulated as the situation is not critical but can be attended to within a certain time period.

The Alarm Fatigue-Aware Notification Model

Figure 3 presents our model designed to support reasoning algorithms that decide on the best approach to notify caregivers to avoid alarm fatigue. The reasoning algorithms, which are the focus of this research, decide the following: (1) whether to aggregate alarms, (2) whether to add a false alarm probability (FAP) label to the notification, and (3) who to notify within the group of caregivers.

Figure 3. An architecture designed for health care systems that support patient monitoring and notification capabilities. MPM: Multi-parametric Monitor; API: Application Programming Interface.



System Architecture

Updating the Anomaly Detection, Alarm-Triggering, and Notification Process Through the Addition of Reasoning

Before presenting the reasoning algorithms, we show, in Figure 4, how the reasoning process interacts with the anomaly detection, alarm-triggering, and notification processes.

Figure 5 is an update of Figure 2 including information related to the reasoning activity.

To deal with the decision-making processes occurring during reasoning, we developed the Reasoner entity that is an instance of our reasoning algorithm. The Reasoner is responsible for managing the entire notification process. A high-level representation of the decision-making processes is shown in Figure 6.

Figure 4. Illustration of the inclusion of the state "Reasoning" (inside the hatched rectangle) that determines when an alarm trigger(s) causes a notification.

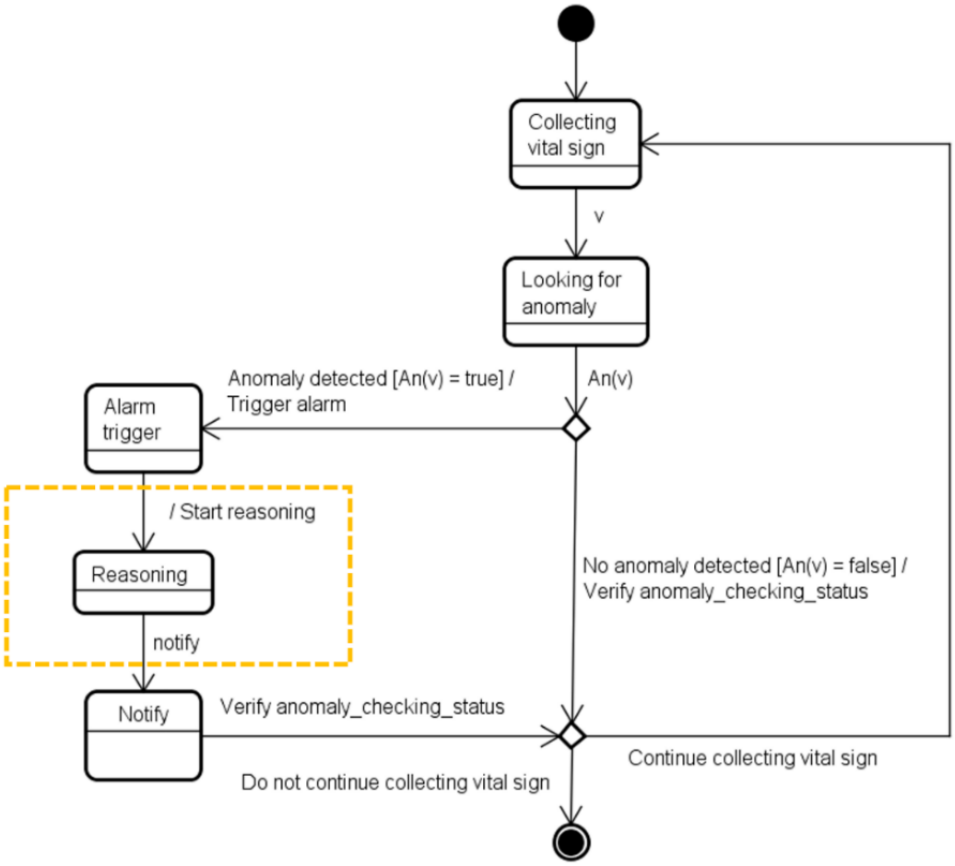


Figure 5. Illustration of the inclusion of the new state "Anomaly alerted under reasoning" (inside the hatched rectangle) as another possible state for an anomaly.

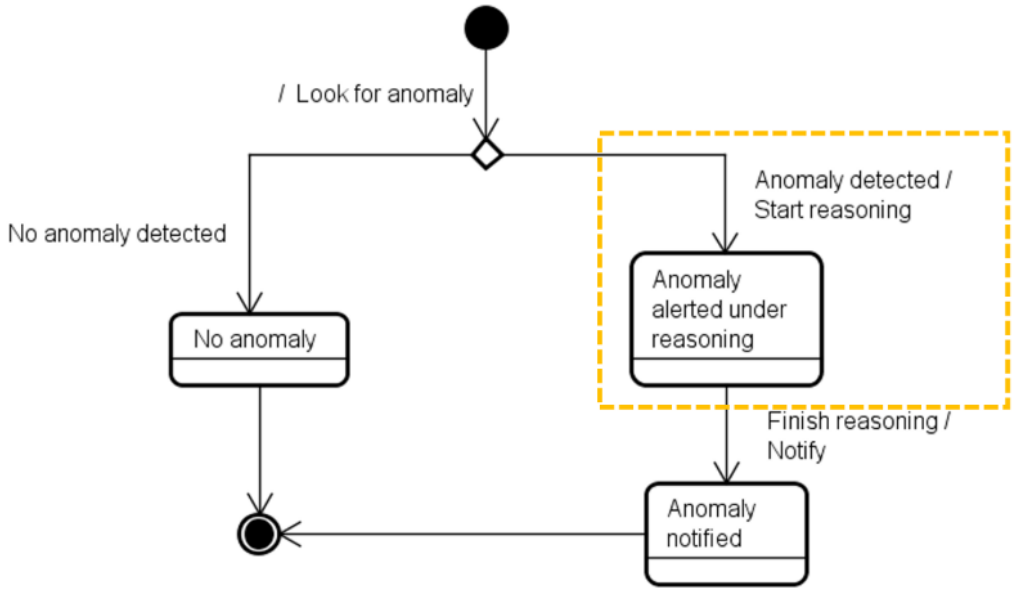
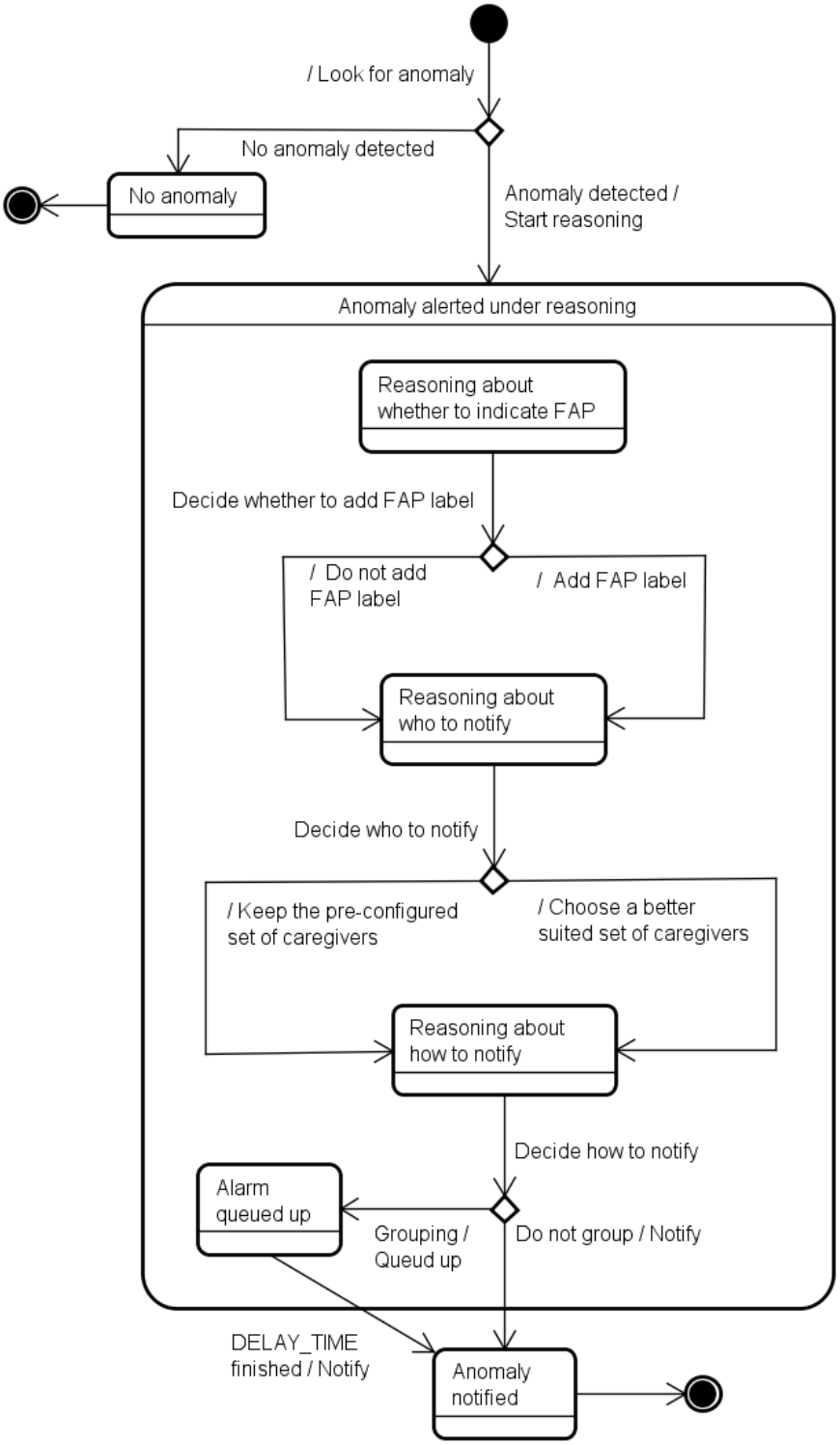


Figure 6. A high-level representation of the decision-making processes used during reasoning. FAP: false alarm probability.



Reasoning About How to Notify to Avoid Alarm Fatigue

In this section, the reasoning algorithm that is used to mitigate alarm fatigue is discussed. As has been mentioned, the default behavior of our anomaly detection process is to trigger an alarm every time an anomaly occurs, independent of circumstances. For example, a notification would occur even though the alarm was false, or a number of other alarms are happening. However, even though an alarm has been triggered by our patient monitoring system, the decision of how to notify the caregivers is decided by the Reasoner, using the following rule R1, which states:

- R1. Our system must limit to one of the number of notifications (of the same type for the same patient) that caregivers can receive within a defined period of time.

MINIMUM_NOTIFICATION_INTERVAL

We define Minimum_Notification_Interval (MNI) as the minimum interval of time between receiving 2 notifications by the caregivers. The R1 rule is only applied when we are considering notifications of the same type (TYPE_β) for the same patient P.

Let μ_j and μ_{j-1} be 2 notifications of the same type for a given patient P. As shown in Equation (9), a notification can be formally defined as μ=<TYPE_μ, T_μ, P>, and in this case we can assume that TYPE_μ_j is equal to TYPE_μ_{j-1} and also that P_j is equal to P_{j-1}. The time T_μ, at which the notification occurs, allows 2 notifications to be distinguished from each other. The MNI can be formally defined in terms of the notifications μ_j and μ_{j-1} as shown below:

$$T_{\mu j} - T_{\mu j-1} >= \text{MNI iff } (\text{TYPE\_}\mu_j = \text{TYPE\_}\mu_{j-1}) \wedge (P_j = P_{j-1}) \quad (10)$$

The MNI value must be configured for each patient individually based on patient's context (both of the alarm sources, and patient's criticality).

The Inputs for Our Reasoning Algorithm Related to a Notification

After explaining rule R1, we define the inputs (*I*) for our algorithm as follows:

- I1—CURRENT_ALARM_TRIGGERING_TIME (Tβr). Let βr be the current alarm that has been triggered and is involved in the reasoning process, so the algorithm can decide whether to add a delay to its delivery. The first input for our algorithm is Tβr, that is, the time when the alarm βr was triggered.
- I2—LAST_NOTIFICATION_TIME (Tμk). Let μk be the last notification (of the same type as βr) received by the caregivers. The second input for our reasoning algorithm is the time when caregivers received μk, which we represent as Tμk.

As we only consider here current alarms under reasoning and last notifications of the same type and from the same patient, we assume that the alarm types and patients are identical, that is, TYPE_β_r=TYPE_μ_k and $P_{\beta r}=P_{\mu k}$.

LAST_NOTIFICATION_PERIOD

Another definition is the Last_Notification_Period (LNP), which is the period of time between the 2 inputs for our reasoning as shown in Equation (11).

$$\text{LNP} = T_{\beta r} - T_{\mu k} \quad (11)$$

The Outputs of Our Reasoning Algorithm Related to a Notification

We next define the outputs (O) for our reasoning algorithm as the 2 properties of notifications that can vary depending on the circumstances under which they occur:

O1—NOTIFICATION_DELAY (D_μ). As discussed previously, in Equation (4), D_μ is the period of time between the *alarm triggering* event and the delivery of that notification to the caregivers.

O2—NOTIFICATION_DATA (DATAμ). DATAμ refers to the type of data a notification might contain, which depends on the context of the alarm-triggering process, and it might range from a single alarm β_j to a set of alarms SET.

As much as possible, we try to keep the NOTIFICATION_DELAY at a minimum so as not to prejudice patient safety. However, to avoid alarm fatigue, the value for this property can range over an acceptable range of time defined as the BUFFERING_PERIOD, indicating that a DELAY_PERIOD (ε) might be added to the delivery time of the notification under specific conditions (defined in the next section). The BUFFERING_PERIOD is the period of time one or more alarms can be delayed (ie, be held in a buffer) before being delivered to caregivers. See Equation (12).

$$0 < \text{BUFFERING\_PERIOD} < \text{MNI} \quad (12)$$

From Equation (12), we show that an alarm might need to be delayed up to a period equal to MNI. However, the BUFFERING_PERIOD specified for an alarm or a set of alarms should not surpass the value of MNI.

Defining the Grouping Criteria for Notification Delivery—When We Shall Put an Alarm Into Our Buffer

As we said previously, the Reasoner decides the way of delivering the alarm under reasoning (β_r) by making choices about whether to add a delay ε to its delivery and whether to group β_r with other alarms. To make these choices, the Reasoner must take into consideration our defined inputs ($T_{\beta r}$ and $T_{\mu k}$). By analyzing these inputs, the Reasoner decides whether to queue the current alarm β_r, based on the following grouping criteria:

- Criteria 1. A same-type alarm was already notified within the MNI.

If caregivers were already notified in the LNP, then the current alarm β_r must be queued up into a buffer for the period BUFFERING_PERIOD. After BUFFERING_PERIOD has passed, β_r is delivered along with other possible alarms in the buffer as a unique notification.

Just to clarify, when the circumstances for the alarms do not meet the abovementioned grouping criteria, a notification containing an individual alarm is sent to the caregivers as soon as an alarm has been triggered, that is, immediately after $T_{\beta r}$.

As important as it is to avoid alarm fatigue, the Reasoner must handle the notification delivery process without compromising patient safety. In this case, the delay added to the notification delivery must not prejudice the requirements established regarding patient safety.

The Pseudocode for Our Reasoning Algorithm About How to Notify

The pseudocode for our reasoning algorithm about how to notify is shown in Textbox 1.

Textbox 1. The pseudocode for the reasoning about how to notify.

```
DEFINE LNP, Tβr, Tμk, MNI;
// Receive Input CURRENT_ALARM_TRIGGERING_TIME Tβr;
INPUT Tβr;
// Receive Input LAST_NOTIFICATION_TIME Tμk;
INPUT Tμk;
// Calculate LNP
LNP=Tβr−Tμk;
// If LNP is equals to Tβr (meaning that no notification μk occurred to the patient in the last MNI-period) or LNP is higher than or equal to MNI (which means that a notification μk occurred more than MNI-period ago) then notify βr immediately. Otherwise, put βr into the buffer
If (LNP==Tβr ||LNP≥MNI) then
//There is no need for putting βr into the buffer. Notify it immediately
Notify(βr);
Else {
// We need to put βr into the buffer and deliver it after some delay
QueuedUp(βr)
// If βr is the first alarm been put into the buffer then {
Ìf (isAlarmTheFirstOneQueuedUp(βr)) then {
//Define buffer's property STARTING_TIME as the time the alarm was triggered;
STARTING_TIME=Tβr;
//Create a new thread for handling the buffer in parallel. This thread needs to
//control the BUFFERING_PERIOD (BP) for notifying caregivers after BP has passed
Create a new thread;
Start BUFFERING_TIME;
If BUFFERING_PERIOD has passed then
//Release the content of buffer to caregivers by wrapping the set of alarms
//(alarmsSet) into a single notification and sending it
Notify(alarmsSet);
}
}
```

Methods

In this work, we present a new approach to cope with the alarm fatigue problem. Our proposed solution focuses on an automatic reasoner that is used to decide how to notify caregivers about anomalies detected by a patient monitoring system through a notification delay strategy.

To confirm the fulfillment of the main research goal, the experiment described next was conducted and results are tabulated in the Discussion section.

Hypotheses

We defined the following hypotheses for our case study:

1. The caregivers should not receive more than one notification about the same type of anomaly for the same patient within the MNI.
2. Patient safety will not be compromised by the use of the reasoning algorithm about how to notify.

Methodology

To illustrate the operation of our reasoning algorithm, we conducted 5 experiments to evaluate how the algorithm works under different scenarios, considering mainly the number of alarms generated in each experiment.

Applications Settings

As shown in Table 1, to run an experiment, we need to define the following settings for our application scenarios:

- The number of wards occupied by patients (NUMBER_OF_WARDS).
- The number of patients being monitored (NUMBER_OF_PATIENTS) by a caregiver team.
- The number of sensors used during monitoring (NUMBER_OF_SENSORS).
- The interval in which the sensor readings are being monitored (SENSORS_READING_INTERVAL).
- The number of sensor readings (NUMBER_OF_READINGS). This information, along with the SENSORS_READING_INTERVAL, tells us how long the patients in our experiment are being monitored.

We also need to define the thresholds for each sensor and the MNI, considering each patient individually (Table 2). As has been mentioned earlier, the MNI is defined by taking into account both of the alarm sources, and the patient's criticality to respect patient safety constraints. In our simulated environment, we defined the MNI value as 5 min for every patient and we assume the delivery of the type of anomalies triggered in our context (which are related to heart rate values) can be delayed up to this period without representing any danger for the patients.

All the inputs for our reasoning were provided through a vital signs streaming app, we developed for streaming vital signs retrieved from a dataset comprising real patient data recorded from patients undergoing anesthesia at the Royal Adelaide Hospital. The dataset provides clinical anesthesia monitoring data from 32 entire surgical cases, including a wide range of vital signs variables, such as electrocardiograph, pulse oximeter, capnograph, noninvasive arterial blood pressure monitor, airway flow, and pressure monitor, and in a few cases, a Y-piece spirometer, an electroencephalogram monitor, and an arterial blood pressure monitor [10]. The monitoring data were collected using Philips IntelliVue MP70 and MP30 patient monitors and Datex-Ohmeda Aestiva/5 anesthesia machines. In this dataset, a single stream of raw monitoring data was recorded in a comma-separated values (CSV) text file format at a sampling resolution of 10 milliseconds [10].

We evaluated our algorithm by using data that we selected from 3 out of the 32 surgical cases in the dataset (cases 04, 07, and 14). Experiment 1 was conducted using data from case 4, while, in experiment 2, we utilized data from case 14, and, finally, experiments 3-5 were executed using data from case 7. In all the experiments, we utilized the version of processed data available in the CSV format for monitoring patients based on their heart rate parameter at 1-second intervals (our algorithm uses this frequency instead of the 10-millisecond sampling resolution available at the dataset). However, the number and type of vital signs used in every experiment could vary to simulate other configurations for sensors and monitoring devices in an ICU.

To define when a given heart rate reading represented an anomalous value that should trigger an alarm, we defined the thresholds in Table 2 for each patient.

Table 1. Defining the configuration for our 5 experiments.

Number of wards	Number of patients	Number of sensors	Sensors reading interval (ms)	Number of readings
1	1	1	1000	60,000

Table 2. Defining the anomaly thresholds of heart rate sensor for each patient.

Experiment	Patient_ID	Min_heart rate	Max_ heart rate
1	1	60	100
2	2	55	100
3	3	50	105
4	4	50	100
5	5	50	102

Results

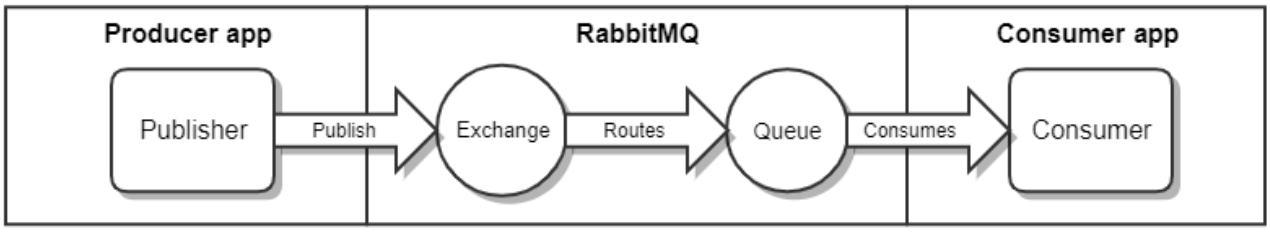
Application Details—Technologies Utilized

The application was developed in the Java language along with the use of the RabbitMQ [13] message broker. RabbitMQ is an open-source message broker that accepts, stores, and forwards messages. The basic concepts behind this technology are Queue, Producer, and Consumer (Figure 7). A Queue is essentially a large message buffer that stores the messages, while a Producer and a Consumer are both user applications. The former is a program in charge of sending messages to the queue through the exchanges, and the latter consists of a program that receives messages from the queue. A program can be both a Producer and a Consumer at the same time [13].

As can be seen from Figure 7, a broker receives messages from publishers (producers) and routes them to the consumers. The information flow involved in this process occurs in 2 steps, described as follows:

- Step 1. The producers send messages to exchanges that act by distributing messages to queues using rules called bindings.
- Step 2. The broker either delivers messages to consumers subscribed to queues or consumes pull messages from queues on demand.

Figure 7. Basic concepts and information flow in RabbitMQ.



In this application, we used the Advanced Message Queuing Protocol 0-9-1 Java client provided by RabbitMQ, which is an open and general-purpose protocol for messaging.

Owing to the high volume of notifications we are dealing with in our application, we decided to utilize a solution that could take care of the nonfunctional requirements of our system. By using a solution to handle problems related to scalability and safety, we could focus on the functional requirements of our application. Therefore, we decided to use the RabbitMQ to meet the high availability, throughput, and scale requirements of our application domain. This message broker solution offers features related to data safety such as reliable delivery, which means it can ensure that messages are always delivered, even encountering failures such as network failures and consumer application failures [13].

Explaining How Our Application Works

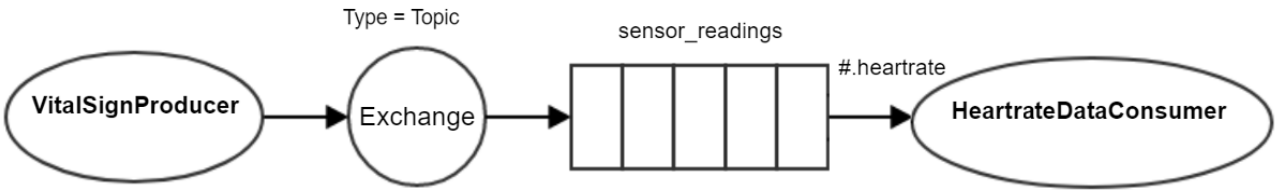
In a high abstraction level, the main idea of this app is to have an application that sends alarms to a broker that routes them to a consumer app that represents the receiving of these alarms by the health care team.

We chose the type of exchange called *topic* for routing the messages. The topic exchange routes messages to one or many queues based on matchings between a message routing key and the pattern that is used to bind a queue to an exchange. We declared one queue named *sensor_readings* to where the publisher sends the data and the consumer receives data. We also declared the binding key for our consumer (ie, the class that is consuming heart rate data) as *#.heartrate* (Figure 8).

The routing key is defined based on the pattern <patientID>.<heartrateValue>. For example, we could have a routing key as *16.88*, representing a patientID=16 and heartrateValue=88.

The notifications sent to health providers are created based on this message. In this case, the final notification received by nurses contains information related to the patient, such as identification, location, and vital signs.

Figure 8. RabbitMQ scheme utilized in our application.



Application Modeling—Class Diagram

In Figure 9, as can be seen from the class diagram for our application, the consumer application monitors a specific vital sign based on the anomalies settings defined for each patient. The consumer app invokes the reasoning mechanism through the ReasoningAboutHowToNotify class, which knows how to notify based on the defined notifications settings (eg, the MNI value configured for each patient).

We present the results of our algorithm by using graphs we generated using the R language and the ggplot2 [14] library. The graphs shown in Figures 10-13 illustrate the delivery process of all notifications related to the patient monitored in experiment 5 (PatientID=5). We show whether the algorithm decided to deliver an alarm immediately or after a delay by grouping alarms to deliver them together.

To better visualize the results of experiment 5 through the graphs, we split the output data of our algorithm for this experiment (comprising a total of 204 alarms) into 4 pieces of data containing 51 alarms each. Thus, we plot each piece of data into a graph, showing the alarm triggering time through the *x*-axis and the notification time on the *y*-axis. As can be seen from Figure 10, the occurrence of the first notification (NotificationID=1) of an alarm of heart rate for patient 5 happened at the notification time *2019-10-01 02:21:41.767*, that is, almost immediately after the occurrence of the first alarm (that happened at the alarm triggering time *2019-10-01 02:21:41.746*). Following the strategy of our reasoning algorithm, the next notification of an alarm of heart rate for this patient should not be received by the caregivers before MNI. As in this experiment MNI corresponds to 5 min, the timestamp for the next delivery of a heart rate alarm related to patient 5 should occur at least 5 min after *2019-10-01 02:21:41.767*. As can be seen in Figure 10, the next heart rate alarms for patient 5 were held in the alarms buffer and delivered together at the timestamp *2019-10-01 02:26:41.77* as a unique notification (NotificationID=2) with a delay of approximately 5 min.

Figure 9. The class diagram for our application, where the consumer application monitors a specific vital sign based on the anomalies settings defined for each patient.

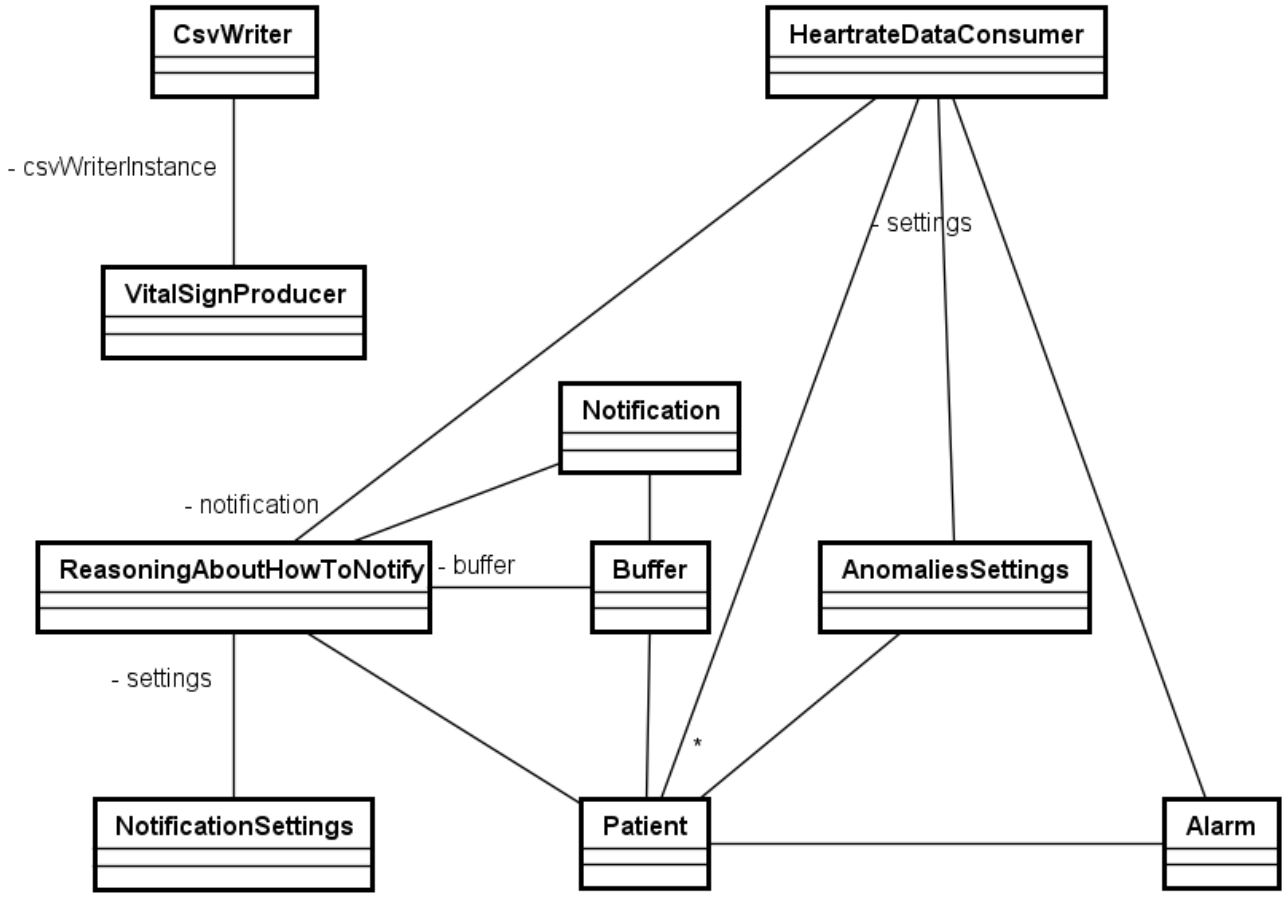


Figure 10. Illustration of the results of the alarm triggering and delivery processes related to the patient monitored in our experiment 5 (PatientID=5).

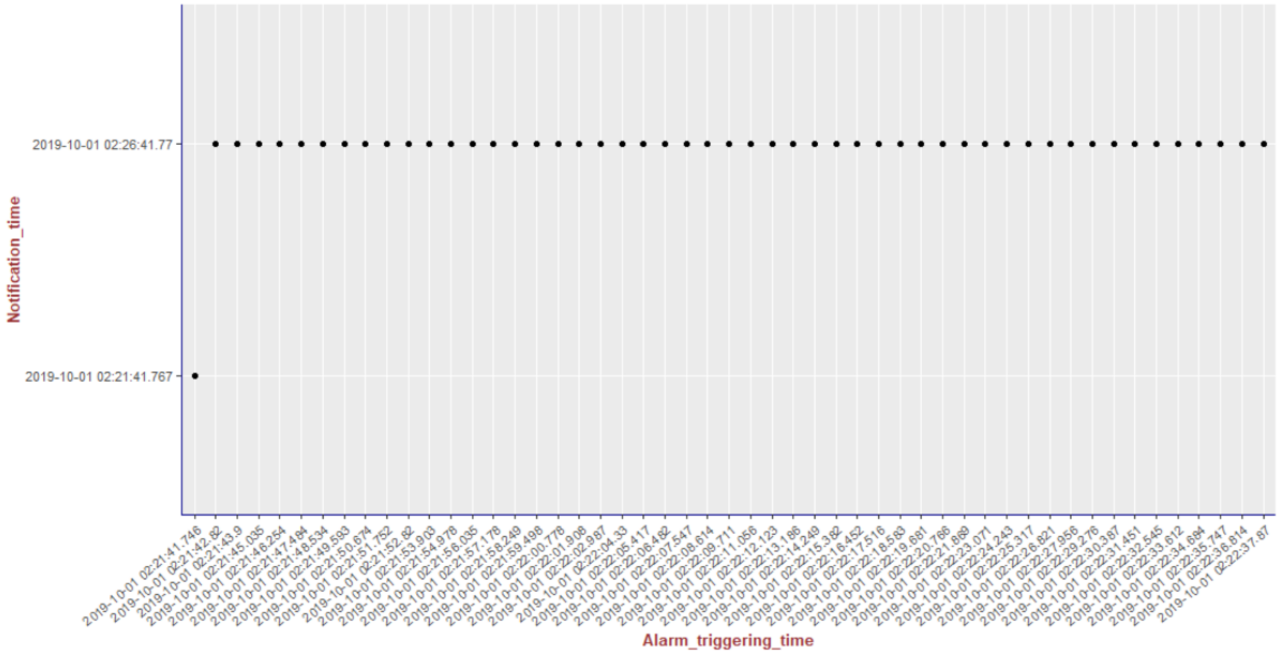


Figure 11. Illustration of the results of the alarm triggering and delivery processes related to the patient monitored in our experiment 5 (PatientID=5).

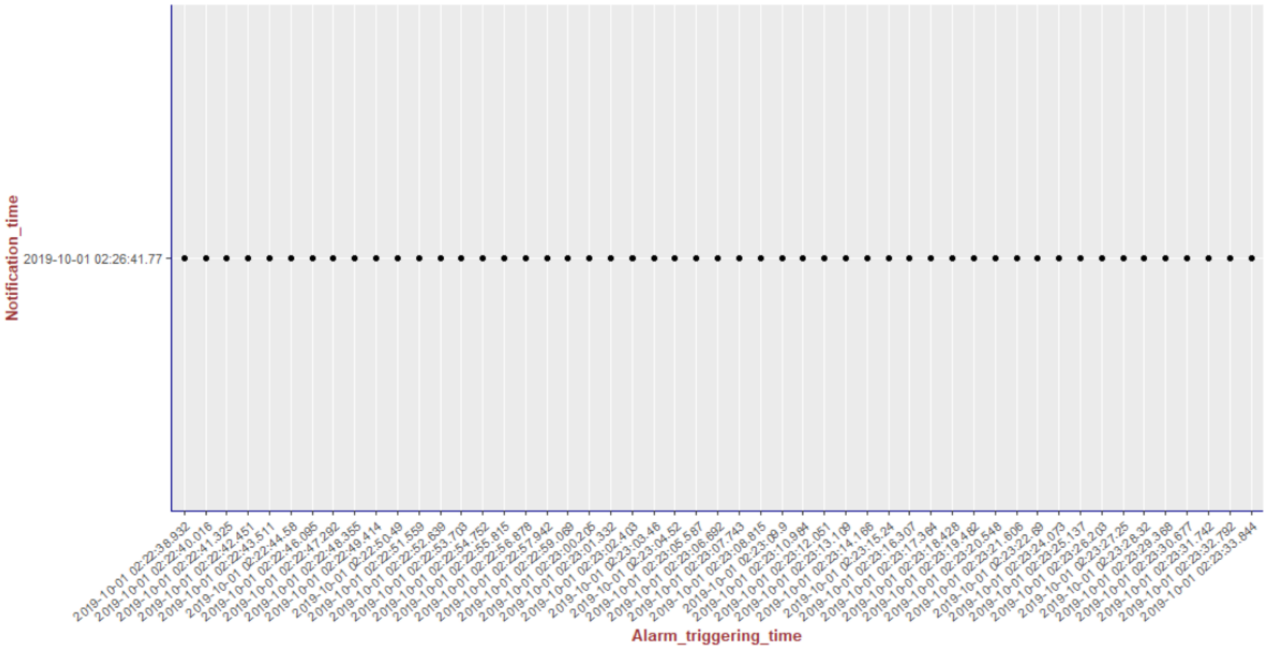


Figure 12. Illustration of the results of the alarm triggering and delivery processes related to the patient monitored in our experiment 5 (PatientID=5).

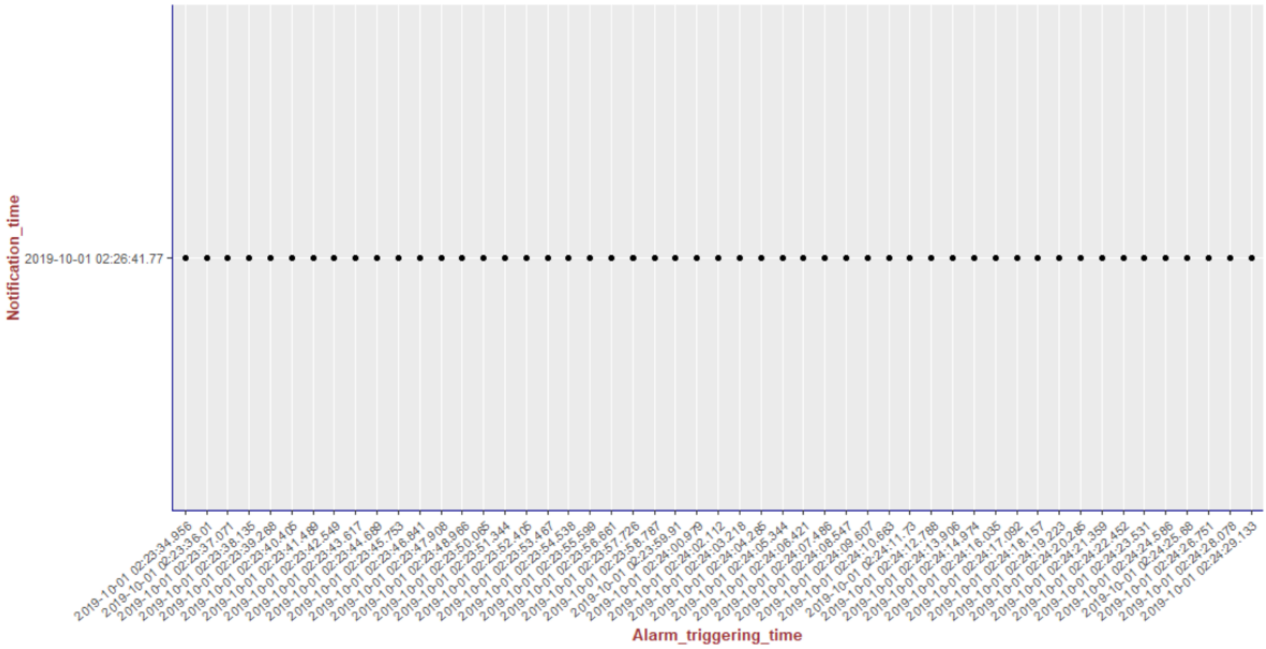
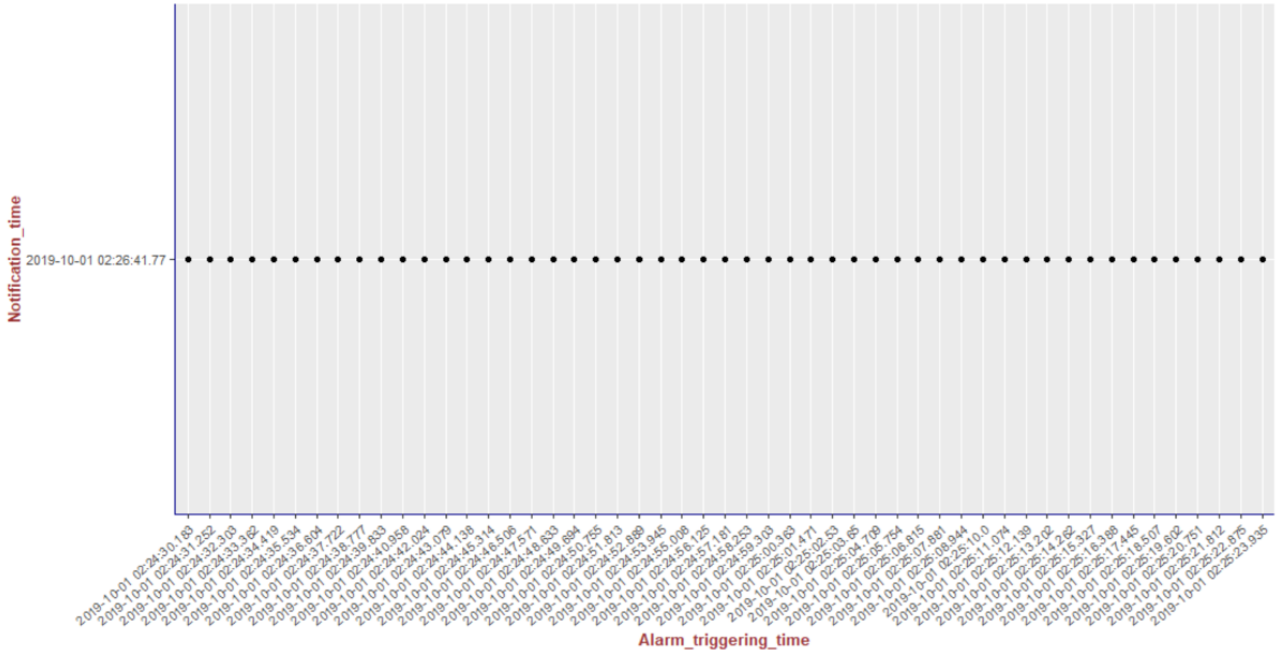


Figure 13. Illustration of the results of the alarm triggering and delivery processes related to the patient monitored in our experiment 5 (PatientID=5).



Figures 14-18 illustrates the results of the delivery processes related to all patients monitored in our experiments (PatientID=1,2,3,4, and 5, respectively).

We show the results for all of our experiments summarized in Table 3, where we can compare the number of alarms triggered by our system in each experiment with the number of notifications delivered to the caregivers.

Figure 14. Illustration of the results of the delivery process related to all our experiments.

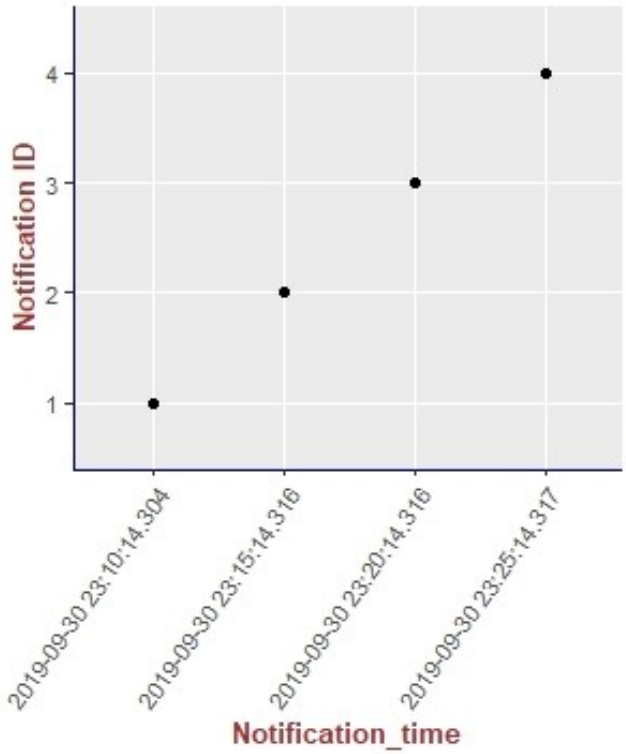


Figure 15. Illustration of the results of the delivery process related to all our experiments.

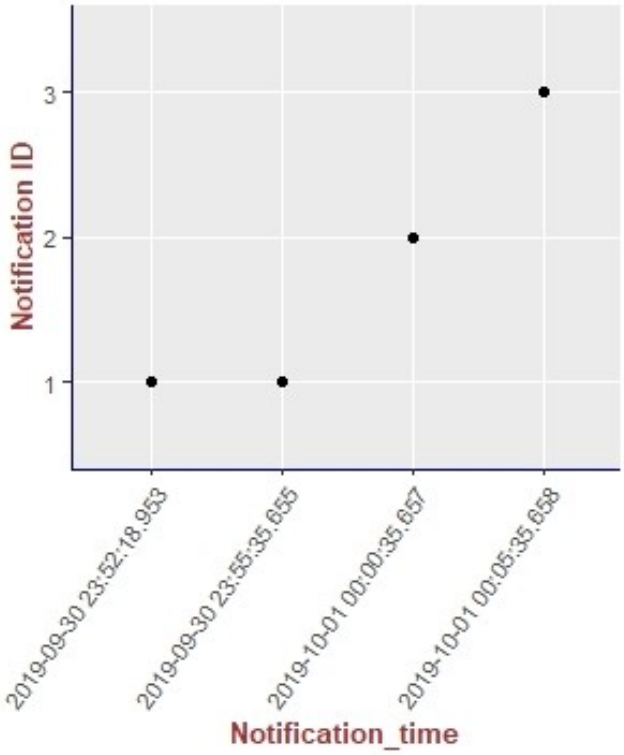


Figure 16. Illustration of the results of the delivery process related to all our experiments.

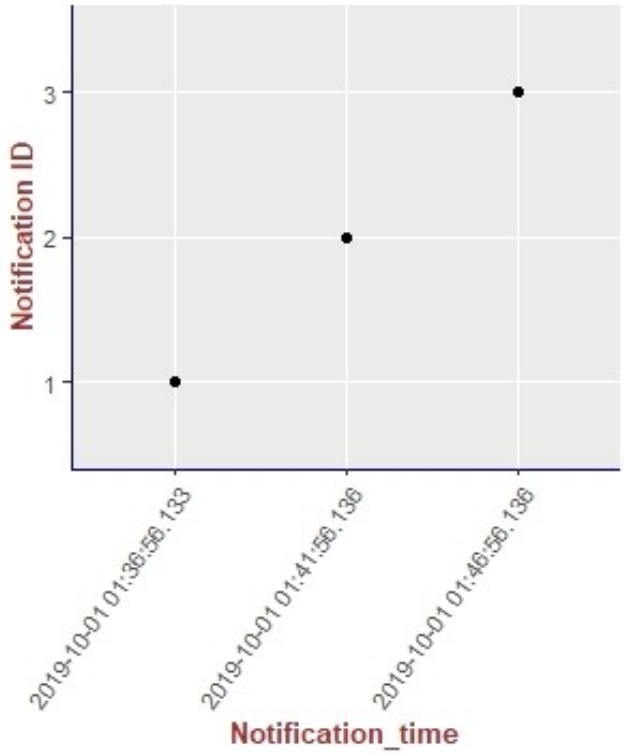


Figure 17. Illustration of the results of the delivery process related to all our experiments.

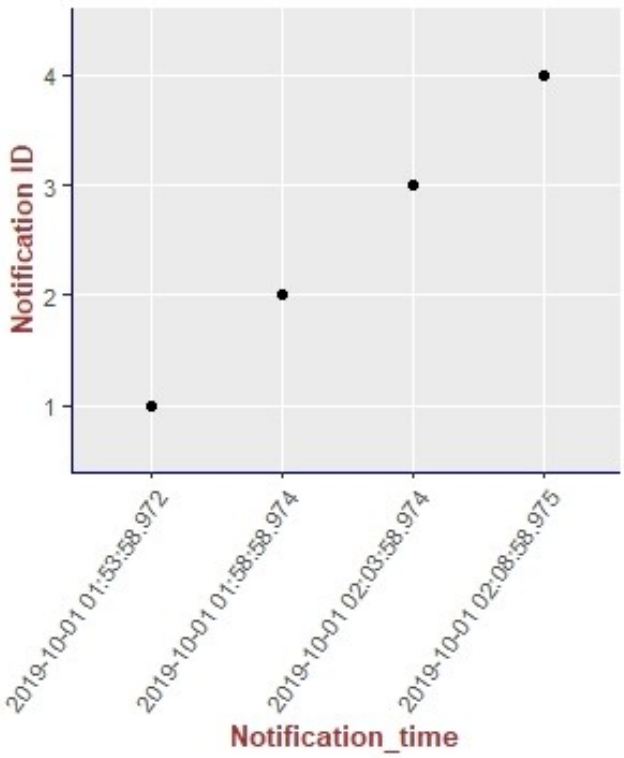


Figure 18. Illustration of the results of the delivery process related to all our experiments.

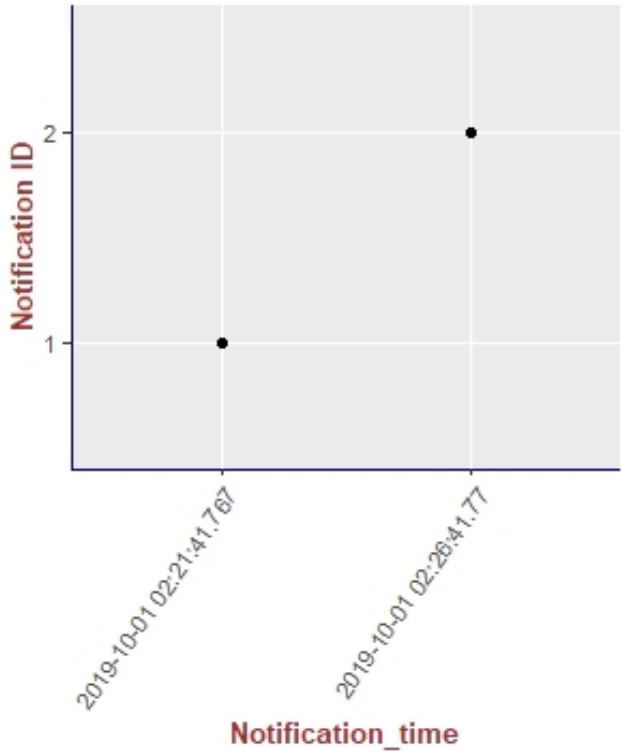


Table 3. Results of our experiments to evaluate our reasoning algorithm about how to notify caregivers considering the reduction of the number of notifications received by them.

Experiment	Heart rate alarms, n	Heart rate notifications,	Notifications in relation to the total of alarms, %	Reduction in alarms received, %
1	407	4	0.9	99.0
2	423	3	0.7	99.2
3	308	3	0.9	99.0
4	586	4	0.6	99.3
5	204	2	0.9	99.0

Discussion

The first hypothesis we want to evaluate with this case study says that the caregivers should not receive more than one notification about the same type of anomaly for the same patient within the defined MNI. By executing our reasoning algorithm throughout the experiments, we saw that hypothesis 1 holds for all of them, as within all the occurrences of notifications for each patient, there is no occurrence of a notification of the same type within the defined MNI. We support this affirmation by presenting, in Figures 14-18, a summary of the results from our experiments using graphs containing all notifications that occurred in each experiment. As can be seen, considering all experiments, there was no occurrence of delivery of notifications of the same type for the same patient that happened before the specified delay, that is, the MNI value of 5 min.

The second hypothesis that says that patient safety will not be compromised by the use of the reasoning algorithm about how to notify also holds, as the notification interval (MNI) we defined is no longer than 5 min. This means that a group of alarms that are happening to a given patient can be held in a buffer for, at most, 5 min before the buffer is fully released to the caregivers as a unique notification. However, in order not to prejudice patient safety, the first occurrence of an alarm is always delivered to the caregivers immediately after its occurrence. In this case, only the next occurrences of the alarms are delivered to caregivers with the addition of a given delay.

In Table 3, we made a comparison between the number of alarms triggered by our system and the number of notifications delivered to the caregivers, in each experiment. These results show that the reduction of the notifications received by the caregivers can be up to 99.3% (582/586) of the total of alarms, with a mean of 99.17% (1912/1928) of reduction in the number of total alarms, considering all the experiments.

According to Winters et al, nearly all studies assume that a reduction in the number of total alarms and/or false alarms will reduce alarm fatigue [9]. Thus, by presenting these results, we expect that our algorithm can be used as a useful strategy for avoiding alert fatigue. We also expect our approach can be useful for helping to prevent its negative consequences, such as disruption of patient care, disabling of alarm systems by staff, reduction in responding, lack of caregiver response, and real events being less likely to be acted on, among others.

In future work, we are planning to extend our approach to reason about whether to notify the caregivers' team with an indication of a FAP. The idea is to provide a reliable classification system in which caregivers may trust so the FAP label added to the notification can help them prioritize their work, especially when they are under alarm fatigue conditions.

Other important future work focuses on how to use reasoning to decide whom to notify within the group of caregivers, considering their specialization level, degree of experience, availability, geolocation, and current workload conditions.

Note that our system is experimental and does not consider security, something that needs to be taken very seriously in an operational health care alarm system.

Acknowledgments

This work was supported by grants from two Brazilian Government Agencies: National Council of Research Development and Foundation for Research Support of the State of Rio de Janeiro.

Conflicts of Interest

None declared.

References

1. Cvach M. Monitor alarm fatigue: an integrative review. Biomed Instrum Technol 2012;46(4):268-277. [doi: 10.2345/0899-8205-46.4.268] [Medline: 22839984]
2. Drew, Harris P, Zègre-Hemsey JK, Mammone T, Schindler D, Salas-Boni R, et al. Insights into the problem of alarm fatigue with physiologic monitor devices: a comprehensive observational study of consecutive intensive care unit patients. PLoS One 2014;9(10):e110274 [FREE Full text] [doi: 10.1371/journal.pone.0110274] [Medline: 25338067]

3. Tanner T. The problem of alarm fatigue. Nurs Womens Health 2013;17(2):153-157. [doi: 10.1111/1751-486X.12025] [Medline: 23594329]
4. Jones K. Alarm fatigue a top patient safety hazard. Can Med Assoc J 2014 Feb 18;186(3):178 [FREE Full text] [doi: 10.1503/cmaj.109-4696] [Medline: 24418978]
5. Shanmugham M, Strawderman L, Babski-Reeves K, Bian L. Alarm-related workload in default and modified alarm settings and the relationship between alarm workload, alarm response rate, and care provider experience: quantification and comparison study. JMIR Hum Factors 2018 Oct 23;5(4):e11704 [FREE Full text] [doi: 10.2196/11704] [Medline: 30355550]
6. Sowan AK, Gomez TM, Tarriela AF, Reed CC, Paper BM. Changes in default alarm settings and standard in-service are insufficient to improve alarm fatigue in an intensive care unit: a pilot project. JMIR Hum Factors 2016 Jan 11;3(1):e1 [FREE Full text] [doi: 10.2196/humanfactors.5098] [Medline: 27036170]
7. Karnik A, Bonafide CP. A framework for reducing alarm fatigue on pediatric inpatient units. Hosp Pediatr 2015 Mar;5(3):160-163 [FREE Full text] [doi: 10.1542/hpeds.2014-0123] [Medline: 25732990]
8. Imhoff M, Kuhls S, Gather U, Fried R. Smart alarms from medical devices in the OR and ICU. Best Pract Res Clin Anaesthesiol 2009 Mar;23(1):39-50. [doi: 10.1016/j.bpa.2008.07.008] [Medline: 19449615]
9. Winters BD, Cvach MM, Bonafide CP, Hu X, Konkani A, O'Connor MF, Society for Critical Care Medicine Alarm and Alert Fatigue Task Force. Technological distractions (part 2): a summary of approaches to manage clinical alarms with intent to reduce alarm fatigue. Crit Care Med 2018 Jan;46(1):130-137. [doi: 10.1097/CCM.0000000000002803] [Medline: 29112077]
10. Liu D, Görges M, Jenkins SA. University of Queensland vital signs dataset: development of an accessible repository of anesthesia patient monitoring data for research. Anesth Analg 2012 Mar;114(3):584-589. [doi: 10.1213/ANE.0b013e318241f7c0] [Medline: 22190558]
11. Fernandes CO, Lucena CJ. A software framework for remote patient monitoring by using multi-agent systems support. JMIR Med Inform 2017 Mar 27;5(1):e9 [FREE Full text] [doi: 10.2196/medinform.6693] [Medline: 28347973]
12. Fernandes CO, de Lucena CJ, de Souza e Silva D. Smart Depth of Anesthesia Monitoring With EEG Sensors and Agent-based Technology. In: 2017 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computed, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation. 2017 Presented at: (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)'17; August 4-8, 2017; San Francisco, CA, USA URL: https://doi.org/10.1109/UIC-ATC.2017.8397455 [doi: 10.1109/UIC-ATC.2017.8397455]
13. RabbitMQ. URL: https://www.rabbitmq.com/ [accessed 2019-06-02] [WebCite Cache ID 78odvPSaX]
14. Hadley W. ggplot2 - Elegant Graphics for Data Analysis. Springer-Verlag New York: Springer; 2009.

Abbreviations

AI: artificial intelligence
AV: anomalous values
FAP: false alarm probability
ICU: intensive care unit
IT: information technology
LNP: last notification period
MNI: minimum notification interval

Edited by G Eysenbach; submitted 08.07.19; peer-reviewed by M Cvach, H Yu, M Görges; comments to author 21.08.19; revised version received 14.10.19; accepted 17.10.19; published 26.11.19.

Please cite as:
Fernandes CO, Miles S, Lucena CJPD, Cowan D
Artificial Intelligence Technologies for Coping with Alarm Fatigue in Hospital Environments Because of Sensory Overload: Algorithm Development and Validation
J Med Internet Res 2019;21(11):e15406
URL: http://www.jmir.org/2019/11/e15406/
doi:10.2196/15406
PMID:31769762

Review

The Personalization of Conversational Agents in Health Care: Systematic Review

Ahmet Baki Kocaballi[1], MSc, PhD; Shlomo Berkovsky[1], PhD; Juan C Quiroz[1], PhD; Liliana Laranjo[1], MD, MPH, PhD; Huong Ly Tong[1], BHealth; Dana Rezazadegan[1], PhD; Agustina Briatore[2], MD; Enrico Coiera[1], MBBS, PhD

[1]Australian Institute of Health Innovation , Faculty of Medicine and Health Sciences, Macquarie University, Sydney, Australia

[2]Health Information Systems Office, Ministry of Health, Buenos Aires, Argentina

Corresponding Author:
Ahmet Baki Kocaballi, MSc, PhD
Australian Institute of Health Innovation
Faculty of Medicine and Health Sciences
Macquarie University
Level 6, 75 Talavera Road
Sydney, 2109
Australia
Phone: 61 298502465
Email: baki.kocaballi@mq.edu.au

Abstract

Background: The personalization of conversational agents with natural language user interfaces is seeing increasing use in health care applications, shaping the content, structure, or purpose of the dialogue between humans and conversational agents.

Objective: The goal of this systematic review was to understand the ways in which personalization has been used with conversational agents in health care and characterize the methods of its implementation.

Methods: We searched on PubMed, Embase, CINAHL, PsycInfo, and ACM Digital Library using a predefined search strategy. The studies were included if they: (1) were primary research studies that focused on consumers, caregivers, or health care professionals; (2) involved a conversational agent with an unconstrained natural language interface; (3) tested the system with human subjects; and (4) implemented personalization features.

Results: The search found 1958 publications. After abstract and full-text screening, 13 studies were included in the review. Common examples of personalized content included feedback, daily health reports, alerts, warnings, and recommendations. The personalization features were implemented without a theoretical framework of customization and with limited evaluation of its impact. While conversational agents with personalization features were reported to improve user satisfaction, user engagement and dialogue quality, the role of personalization in improving health outcomes was not assessed directly.

Conclusions: Most of the studies in our review implemented the personalization features without theoretical or evidence-based support for them and did not leverage the recent developments in other domains of personalization. Future research could incorporate personalization as a distinct design factor with a more careful consideration of its impact on health outcomes and its implications on patient safety, privacy, and decision-making.

(J Med Internet Res 2019;21(11):e15360) doi:10.2196/15360

KEYWORDS

conversational interfaces; conversational agents; dialogue systems; personalization; customization; adaptive systems; health care

Introduction

Background

Recent advancements in natural language recognition and synthesis have resulted in the adoption of conversational agents (CAs) in many fields. CAs can be defined as systems that support conversational interaction with users by means of speech or other modalities [1]. The rising popularity of conversational technologies has been facilitated by a renaissance in Artificial Intelligence, the development of powerful processors supporting deep learning algorithms, and technological advancements, making a large amount of computationally accessible knowledge available [1].

One emerging area in which conversational technologies have been increasingly used is health care. A recent systematic review in this area examined technical performance, user experience, and health-related outcomes and found that most studies had not employed standardized evaluation methods or had failed to address aspects of patient safety [2]. There have also been other recent reviews on health care conversational agents [3-5]. This study differs from them in that it focuses on the implementation of personalization in health care conversational agents.

Personalization

Personalization is:

> *the process of making something suitable for the needs of a particular person [6].*

When applied specifically to digital technologies, personalization can be defined as:

> *a process that changes the functionality, interface, information access and content, or distinctiveness of a system to increase its personal relevance to an individual or a category of individuals [7].*

A recent interdisciplinary review study proposed a framework to characterize personalization along three dimensions: (1) what is personalized (ie, content, user interface, delivery channel, and functionality); (2) for whom is it personalized (either a specific individual or a user group, eg, elderly women); and (3) how automated is the personalization (how the information needed for user modelling is collected) [7]. The personalization process involves user models containing characteristics, preferences, interests, and needs of users as the basis for providing adaptive information and services. Depending on the degree of automation, two types of personalization can be distinguished: implicit and explicit. In implicit personalization, information needed for user models is obtained automatically through the analysis of observed user activities and interactions with the system. In explicit personalization, information needed for user models requires users' active participation in obtaining the required information.

Personalization in Conversational Agents

One of the earliest applications of personalization in a conversational system was Grundy, a virtual librarian that delivered book recommendations [8]. To build a user model for personalization purposes, Grundy asked questions at the beginning of an interaction and associated users with predefined stereotypes. After the initial user provided information, the user model was updated implicitly over time during conversations. It was a hybrid system bringing together both explicit and implicit personalization. This foundational work on personalized CAs has been followed by a range of works focusing on dialogue management [9], personalized messages [10], recommender systems [11], and adaptive systems [12].

Personalization in CAs can be achieved implicitly by processing past interactions with users [11,13] or explicitly by user-entered information at the set-up time [8] or using ongoing confirmation style input [14]. The messages presented to users [10], or the conversational style of systems [15], can be personalized. Personalized and adaptive system behavior in conversational systems can improve user comprehension [16], user satisfaction [17], task efficiency [18], and the likelihood of behavior change [19]. Furthermore, personalization can be an essential system feature for voice interfaces due to the limitations in presenting large amounts of information through a voice-only modality [20]. The effects of personalization have been evaluated in various ways by measuring aspects like efficiency in terms of the number and duration of interactions [11,20], user satisfaction, relevance, and understandability [20], information quality presented [21], and appropriateness of system responses [10].

Personalization in Health Care and Medicine

Studies of personalization in health care and medicine have been increasing in number since the early 2000s [22], with growing evidence showing their effectiveness [23-26]. One important limitation in the health care personalization literature is equating it to genomics-supported efforts in medicine [27]. Genomic markers are only one dimension of personalization that helps to recognize the uniqueness of individuals and make their medicine personalized [27,28]. There are other factors that affect this personalization of health care, such as people's lifestyle choices, their socioeconomic context and living environment, and other health care services that can be personalized like health education and therapies [29].

A review of behavior change interventions characterized four intervention groups according to their degree of personalization in the messages delivered to individuals: generic (one-size-fits-all messages), personalized (messages with the person's name), targeted (messages specific to a subgroup of the general population), or tailored (messages specific to an individual's characteristics) [30]. The review found that 78% (11/14) of the tailored and 95% (22/23) of the targeted interventions reported improved outcomes, with 54% (6/11) of the tailored and 68% (15/22) of the targeted interventions being statistically significant.

Dialogue systems can offer fine-grained possibilities to personalize the information to be delivered:

> *on the basis of the inferred goals and beliefs of the user at a particular moment in time, and incorporating everything that has previously been said in the conversation [31].*

Learning from a history of previous conversations plays a key role in ensuring the continuity of health communications that take place over multiple interactions over time [31].

Informed by the recent theoretical developments in personalization [7], a broader understanding of personalization in health care [22,29], and an increasing interest in health care CAs [3-5], this study aims to review the use of personalization in health care CAs and characterize the methods that have been applied to implement this personalization. Aligned with the rapid advancements in natural language processing technologies used in CAs [1] and the increasing adoption of CAs using unconstrained natural language [32], this review focuses on agents with unconstrained natural language input capability. These agents include chatbots, which can engage in small talk or casual dialogues, embodied conversational agents, which

feature computer-generated visual virtual characters capable of both verbal and nonverbal communication, and commonly available smart conversational interfaces such as Apple's Siri, Google's Google Assistant, Samsung's Bixby, and Microsoft's Cortana [1,33,34].

Methods

Overview

This review uses the search protocol of an earlier systematic review that was performed between April 2017 and February 2018, with a focus on technical performance, user experience, and the health-related outcomes of CAs in health care [2]. The current review has: (1) focused on the use of personalization features in CAs that were not examined previously; (2) used the same inclusion and exclusion criteria as the review by Laranjo et al [2] with an additional criterion on personalization (ie, the studies with no personalization features were excluded); and (3) performed a new search in March 2019.

Search Strategy

We searched in the PubMed, Embase, CINAHL, PsycInfo, and the ACM Digital Library databases, and did not restrict by the publication year or language. The search terms included "conversational agents", "dialogue systems", "relational agents" and "chatbots". The complete search strategy is available in Multimedia Appendix 1. In addition, the reference lists of relevant articles and grey literature identified in those databases were also included for screening.

Study Selection Criteria

The identified publications were included if they: (1) were primary research studies that focused on consumers, caregivers, or health care professionals; (2) involved a conversational agent; and (3) tested the system with human users. The studies were excluded if they involved: (1) user input by clicking or tapping an answer amongst a set of predefined choices, or by using the telephone keypad (eg, interactive voice response systems with dual tone multi frequency); (2) output not generated in response to what it received from the human user (eg, predefined and preprogrammed messages that are not dependent on the information obtained from or about the user); (3) question-answer type interactions; (4) asynchronous communication technology such as email; or (5) no personalization features. Furthermore, studies evaluating only individual components of a conversational agent, like automatic speech recognition, or using Wizard of Oz methods were excluded.

Screening, Data Extraction, and Synthesis

Screening was conducted independently by two researchers to extract data from each study. Cohen kappa was used to measure intercoder agreement between the researchers. Any disagreements between the assessments of two researchers were resolved by consensus agreement. To identify the relevant information, the researchers used the personalization definition presented in the introduction section. In addition, the following keywords were used as a guide to identify personalization-related information within the studies: personalizing, adapting, customizing, tailoring, configuring, individualizing, modifying, changing, altering, transforming, modelling, tuning, setting, preference, and profile. The data extraction process was guided by an assessment scheme based on the personalization framework offered by Fan and Poole [7]. In addition to these dimensions, we included three more dimensions to provide further details on the included studies: purpose of personalization, methods to evaluate personalization, and outcomes in relation to personalization. Table 1 summarizes the final assessment scheme for personalization.

Table 1. An assessment scheme for personalization.

Assessment categories	Description
Automation[a]	How the user models needed by personalization are constructed.
Implicit	Information needed for user models is obtained automatically through the analysis of observed user activities and interactions with the system (eg, analyzing users' conversational history to determine the suitable times to send a reminder).
Explicit	Information needed for user models requires users' active participation in obtaining the required information (eg, selecting the preferred times to receive a reminder).
Target[a]	For whom to personalize.
Individuated	Personalization is targeted at a specific individual (eg, sending a reminder based on the unique profile of a single user).
Categorical	Personalization is targeted at a group of people (eg, sending a reminder based on a shared profile of a group of users).
Aspects of the system[a]	What to personalize.
Content	The information itself (eg, alerts or reminders).
User interface	How the information is presented (eg, using larger font sizes for elderly users or shortening prompts for experienced users).
Delivery channel	The media through which information is delivered (eg, sending a reminder as a text message instead of a voice message).
Functionality	What users can do with the system (eg, making different system functionalities available for patients and carers).
Purpose	The purpose of personalization (eg, increasing user engagement or motivation).
Evaluation	The methods to evaluate personalization (eg, using interview questions or standardized questionnaires).
Outcomes	The outcomes in relation to personalization (eg, increased user engagement or motivation).

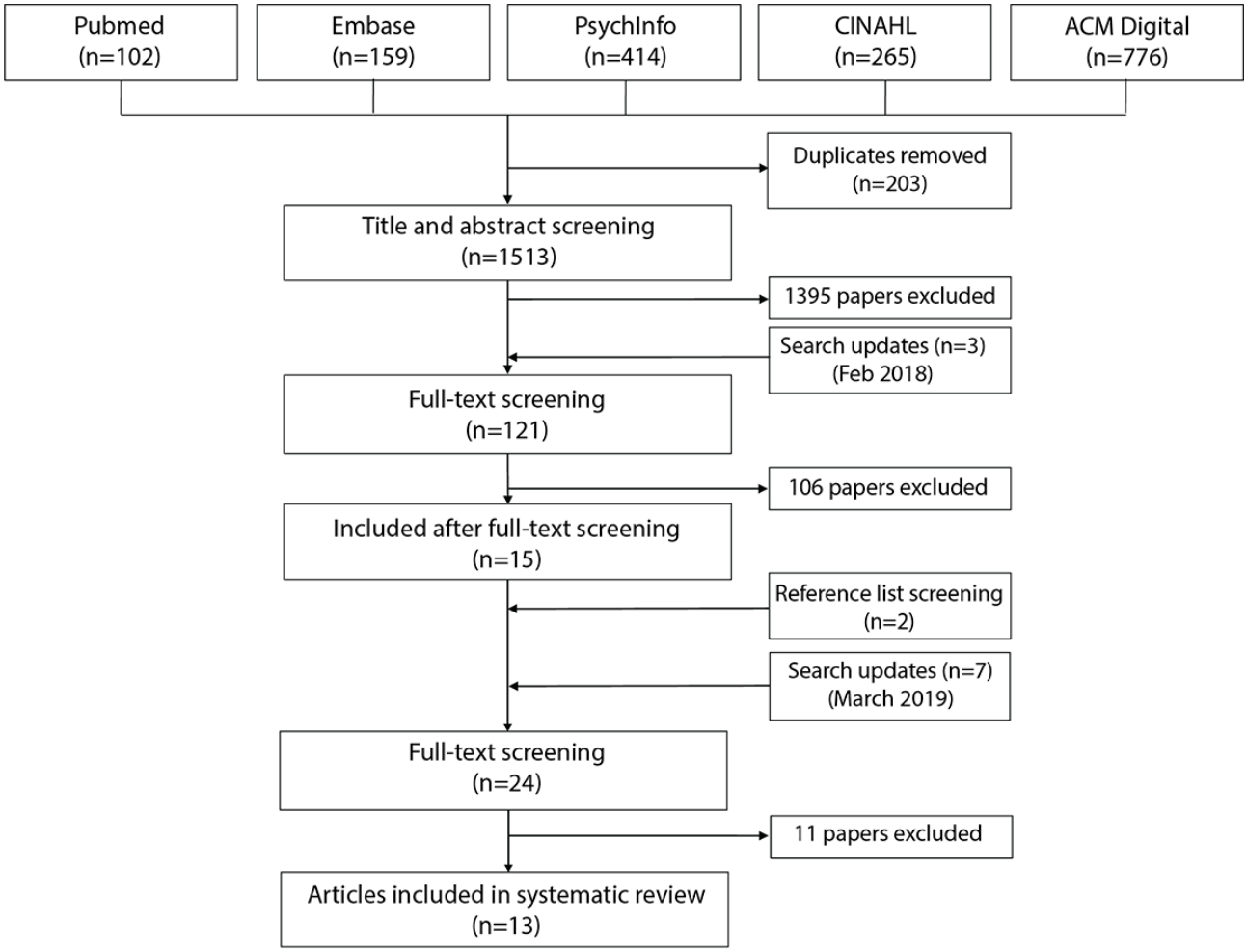
[a]Adapted from Fan and Poole [7].

Results

Search Results

The first search found 1513 papers, and the updated search found an additional 445 papers (Figure 1). After the subsequent title, abstract and full text screenings, 13 studies were included in this review [35-47]. The first search's kappa statistic for the title and abstract screening was 0.45 (fair agreement) and for the full-text screening it was 0.53 (fair agreement). For the updated search, the kappa score was 0.77 for the title and abstract screening (substantial agreement), and 0.61 for the full-text screening (substantial agreement). The list of excluded studies, their major themes, and the reasons for their exclusion are available in Multimedia Appendix 2.

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow diagram.



Implementation of Personalization Features and Target Population

Table 2 and Table 3 summarize the personalization features of CAs in the included studies [35-47]. For both tables, studies evaluating the same conversational agent were grouped together. Since the delivery channel and functionality were not personalized by any of the studies, they were not included in Table 2. Out of the 13, 8 studies supported patients and 5 studies supported both patients and clinicians. Regarding the target of personalization, all the studies implemented individuated personalization (targeting an individual user). However, one study employed categorical personalization (targeting a group of people) to differentiate novice and expert users in addition to the individuated one [41].

Automation of Personalization

Information needed for personalization was provided explicitly by the users in seven studies [35,37-40,44,45], and obtained implicitly by the system in one study [36] where the conversational agent analyzed users' audio-visual features, such as facial expression and head position, to determine its feedback. A mix of implicit and explicit methods was employed by five studies [41-43,46,47]. Across all the studies, data explicitly entered by the users included personal goals [35,37,46,47], symptoms and medications [37,44,45], measurement of vital signs [39,40], knowledge level on a specific topic [38], and daily practices [38]. User data implicitly obtained by the systems primarily involved the analysis of conversation history. Differently from the rest, one study analyzed users' voice and nonverbal facial gestures to determine narrative skills of the users [36].

Table 2. Personalization features of conversational agents in the included studies.

Conversational agent (author, year)	CA[a] purpose	Automation (the basis for personalization)	Target (for whom to personalize)	What to personalize	
				Content	User interface
Tess (Fulmer et al, 2018) [44]	Delivery of cognitive behavioral therapy to reduce symptoms of depression and anxiety in college students	• Explicit: Expressed emotions and mental health concerns of participants to provide personalized responses. Users' feedback and reported mood used to tailor interventions	• Individuated	• Personalized conversations based on emotions and mental health concerns • Personalized therapeutic choices based on user feedback	NR[b]
Wysa (Inkster et al, 2018) [45]	Wellbeing support app for users with symptoms of depression, aiming to build mental resilience and promote mental wellbeing	• Explicit: User responses to built-in assessment questionnaire and emotions expressed in a written conversation	• Individuated	• Personalized conversational pathways based on a user's interaction, messages, and context	NR
Reflection Companion (Kocielnik et al, 2018) [46]	Support reflection on personal physical activity data from fitness trackers	• Explicit: Users enter their behavior change goals and demographic data • Implicit: Observed physical activity of the user	• Individuated	• Dialogues to encourage reflection • Incorporating user goals into adaptive mini-dialogues • Follow-up questions based on users' earlier responses • Visualization of past physical activity	NR
Relational Agent (Sillice et al, 2018) [47]	Promote regular exercising and sun protection	• Explicit: Users provide their demographic information, exercising habits, sun protection behaviors and lifestyle goals • Implicit: CA tracks user progress to send reminders if needed	• Individuated	• Acknowledgement of difficulties and tailored strategies to overcome these • Feedback on progress and encouragement for achieving goals • A weekly tracking chart to help participants monitor their exercise and sun protection behaviors • Email reminders to support retention	NR
Woebot (Fitzpatrick et al, 2017) [35]	Deliver cognitive-behavioral therapy for anxiety and depression to college students	• Explicit: Users enter their mood and goals	• Individuated	• Empathic responses tailored to the reported mood • Tailoring of support content depending on the reported mood • Daily prompting messages to initiate a conversation • Weekly charts depicting the reported mood and textual summary	NR
Social Skills Trainer (Tanaka et al, 2017) [36]	Social skills training for people with autism spectrum disorders	• Implicit: CA analyzes the user's audio-visual features, facial expression (smile), and head position to determine its feedback and then performs feature selection	• Individuated	• Personalized score showing similarity to a role model with respect to 10 features • Encouraging comments to reinforce motivation, based on features closest to the model • Comments on the points that need improvement, based on features dissimilar to the model • Homework challenges for participants to complete on their own time throughout the week	NR

Conversational agent (author, year)	CA[a] purpose	Automation (the basis for personalization)	Target (for whom to personalize)	What to personalize	
				Content	User interface
mASMAA[c] (Rhee et al, 2014) [37]	Facilitate asthma symptom monitoring, treatment adherence, and adolescent-parent partnership	• Explicit: Users enter symptoms, activity level, and use of rescue and control medications	• Individuated	• Automated inquiries and reminders sent according to user-defined preferences on monitoring symptoms and managing medications and activity • Processing of and responses to user-initiated messages at any time • Daily report summarizing symptoms, activity, and use emailed to parents	NR
Chris (Hudlicka, 2013) [38]	Embodied CA that provides mindfulness training and coaching	• Explicit: Users answer questions asked by the CA and set preferences via multiple-choice questions	• Individuated	• CA's facial expressions and its responses adapting to the users' learning needs and motivational state • CA's affective reaction adapting to the users' utterances • Conversational expressions communicating mental state • Customized advice about meditation practice, based on the expressed concerns	Using didactic, relational, or motivational conversational styles according to the user models
DI@l-log (Harper et al, 2008; Black et al, 2005) [39,40]	Voice logbook to document home monitored data by diabetes patients	• Explicit: Users provide weight, blood sugar and blood pressure values	• Individuated	• An alert feature generating a verbal warning if readings are too high • Personalized feedback to patients on their current progress	NR
Pain Monitoring Voice Diary (Levin and Levin, 2006) [41]	Real-time collection of information from patients for health, behavioral, and lifestyle studies and monitoring	• Explicit: Users answer a series of questions about their pain (location, type, intensity, etc) • Implicit: CA utilizes previous sessions to provide personalized content and conversational style	• Individuated • Categorical (novice and experienced users)	• Content (what data is collected) and style (how it is collected) of the reporting session • Adaptive question-asking (additional questions for follow-ups to sessions with high levels of pain) • Adaptive interruptions to better support experienced users	Adaptive conversational style (eg, shorter question formats for follow-up sessions)
Intelligent dialogue system (Giorgino et al, 2004; Azzini et al, 2003) [42,43]	Home care and data acquisition from hypertension patients	• Explicit: Users answer questions about heart rate, pressure, weight, compliance, and more • Implicit: CA changes its behavior depending on the progress of the current call and the clinical history of the caller	• Individuated	• The questions to be asked were determined by user profiles • Gives advice on recommended health behavior and next visits • Issues alerts and prompts	NR

[a]CA: conversational agent.

[b]NR: not reported.

[c]mASMAA: mobile phone-based asthma self-management aid.

Table 3. Personalization purpose, evaluation, and outcomes in the included studies.

Conversational agent (author, year)	Personalization		
	Purpose	Evaluation	Outcomes
Tess (Fulmer et al, 2018) [44]	• To improve depression and anxiety symptoms • To provide more engaging and convenient user experience • To provide appropriate response and strategies based on the users' reported emotion and health concerns	• Questionnaires to measure depression (PHQ-9[a]) [48], anxiety (GAD-7[b]) [49], and affect (PANAS[c]) [50] • Custom-built user satisfaction questionnaire • Number of messages to measure user engagement	• Significantly lower depression (P=.03) and anxiety scores (Group 1, P=.045; Group 2, P=.03) and significant differences in the positive and negative affects (P=.03; sm[d]) • 86% (43/50) of participants satisfied with CA[e] (sm) • Comparable levels of daily engagement (bm[f])
Wysa (Inkster et al, 2018) [45]	• To develop positive self-expression and create a responsive self-reflection environment • To encourage users to build emotional resilience skills	• Questionnaire to measure depression (PHQ-9) • Thematic analysis of the responses to the in-app feedback questions • User engagement through analysis of raised objections and thematic analysis of in-app feedback	• Significant reduction in depression scores in both high (P<.001) and low user groups (P=.01; sm) • 67% (191/282) of users reporting on positive app experience (sm) • More than 99% (6555/6611) of detected objections were correct (bm)
Reflection Companion (Kocielnik et al, 2018) [46]	• To trigger deeper reflection, which would increase motivation, empowerment, and adoption of new behavior • To provide engaging, novel, and diverse conversations around reflection	• Questionnaires to measure health awareness [51], mindfulness (FMI[g]) [52], and reflection (RQ[h]) [53] • Willingness to use the system, number, and length of responses as measures of engagement • Responses to mini-dialogues • Semi-structured post-study interviews	• Significant increases in habitual action (P=.05) and understanding (P=.07; sm) • Prolonged use of CA (additional two weeks) by half of the participants (16/33) with an avg of 98.4-character response length in this period (bm) • High response rates: 96% (443/462) of initial and 90% (386/429) of follow-up questions (bm) • Mini-dialogues successfully supporting discussions on awareness related to goal accomplishment, self-tracking data, and trends in behaviour (nq[i], sm) • Interviews indicating an increase in awareness, mindfulness, and motivation; understanding of alternatives and actions; and newly discovered insights (sm)
Relational Agent (Sillice et al, 2018) [47]	• To increase user engagement and promote more effective behavior change • To monitor exercise and sun protection behavior • To provide strategies to overcome the reported barriers	• Interviews to assess user experience and a 10-point Likert scale to measure satisfaction with interventions	• The levels of satisfaction ranged between 7 and 10 on a scale of 1 to 10 (sm) • Most participants reporting on: (1) positive interactions with the CA (32/34; 94%); (2) tailored feedback supporting regular exercising and sun protection behaviors (29/34; 85%); and (3) email reminders helping to remain on track with the program (23/34; 68%; sm)
Woebot (Fitzpatrick et al, 2017) [35]	• To engage individuals with CA through managing conversation tailored to the reported mood	• Questionnaires to measure depression (PHQ-9), anxiety (GAD-7), and affect (PANAS) • Custom-built questionnaire to measure user satisfaction, emotional awareness, learning, and relevancy of content	• Significant reduction in depression symptoms (P=.04; sm) • Significantly high level of overall satisfaction (P<.001) and greater amount of emotional awareness (P=.02; sm)
Social Skills Trainer (Tanaka et al, 2017) [36]	• To provide personalized feedback aimed at improving narrative social skills	• Experienced human social skills trainer assessed the participants' narrative skills	• Improvements in the overall narrative and social skills (Study 1, P=.03; Study 2, P=.003; bm)

Conversational agent (author, year)	Personalization		
	Purpose	Evaluation	Outcomes
mASMAA[j] (Rhee et al, 2014) [37]	• To make the system more appealing and elicit greater and longer interest in and use of the system	• Six routine asthma-diary questions • Focus group interviews to evaluate user experience with CA	• Improved self-management, treatment adherence, accessibility of advice, awareness of symptoms, and sense of control (nq, sm) • CA was found to be easy-to-use, convenient, and appealing (nq, sm)
Chris (Hudlicka, 2013) [38]	• To deepen the relationship with the user • To support pedagogical strategies necessary for effective training of mindfulness meditation • To provide the coaching required to initiate and maintain regular practice • To provide interactions for maintaining motivation via empathic dialogue and customized advice	• Custom-built questionnaires to assess the overall experience, meditation frequency, knowledge of mindfulness, sense of self-efficacy, and stages of change within the transtheoretical model of change	• Improved outcomes with CA group compared to a self-administered program: (1) more frequent and longer mindfulness training sessions (P=.01); (2) more rewarding, enjoyable, beneficial, and engaging experience (nq); and(3) more advanced stages of change and more confidence in ability to maintain regular meditation (nq) • Neutral to mildly positive feedback on CA's ability to provide customized feedback (0.3 on a –2 to +2 Likert scale; sm)
DI@l-log (Harper et al, 2008; Black et al, 2005) [39,40]	• To provide personalized feedback on the patient's health status and increase their engagement	• Task completion rate and time • Number of personalized alerts • Qualitative interviews	• 92.2% (190/206) successfully completed calls, shortening calls over time, and effective alerts leading to 12 therapeutic interventions (bm) [39] • 90.4% (38/42) successfully completed calls, users' appreciation of the personalization and reports on empowerment, peace-of-mind, and sense of care (bm, sm) [40]
Pain Monitoring Voice Diary (Levin and Levin, 2006) [41]	• To shorten the dialogue sessions • To provide the users a feeling of continuity • To have flexible and adaptive support for different types of users	• Session length, completion rate, and turn duration • Ratio of prompt interruptions by users	• 97% (171/177) of sessions completed with 98% (849/859) input accuracy (bm) • Shortening dialogues over time (avg 1.2 seconds over 7 sessions; bm) • More prompt-interruptions by the experienced users (73% of the prompts) compared to the novice users (59% of the prompts; bm)
Intelligent dialogue system (Giorgino et al, 2004; Azzini et al, 2003) [42,43]	• To improve the quality of system dialogues • To increase patient compliance with guidelines	• Reliability and recognition error rate • Time spent in learning to use the system	• Recognition rate up to 41%-81% (bm) • Dialogue time of 3.3-5.9 minutes, with 80% (74/93) of the expert users' dialogues achieving conclusion (bm)

[a]PHQ-9: Patient Health Questionnaire 9-item scale.

[b]GAD-7: Generalized Anxiety Disorder 7-item scale.

[c]PANAS: positive and negative affect schedule 20-item scale.

[d]sm: self-reported measure.

[e]CA: conversational agent.

[f]bm: behavioral measure.

[g]FMI: Freiburg Mindfulness Inventory.

[h]RQ; Reflection Questionnaire.

[i]nq: not quantified.

[j]mASMAA: mobile phone-based asthma self-management aid.

What is Personalized?

Personalization was primarily used for tailoring the content to be delivered. Personalized content included: (1) feedback on mood states [35], narrative skills [36], symptom summaries [37], meditation practice [38], and current progress towards the goals set [39,40,46,47]; (2) reminders [37,47], warnings, and alerts [39,40,42,43]; (3) multimedia [35,46]; and (4) questions on pain [41], physical activity [46], and health status [42,43].

Two studies personalized the user interface through changing conversational styles according to users' motivation state, users' level of expertise with the system, and dialogue history [38,41]. For example, one study used either didactic, relational, or motivational conversational styles based on the user profile and

progress [38]. While the didactic style was used for training-related conversations, the relational style was used at the beginning of sessions to improve user engagement based on the answers received from the user. The motivational style was employed to gather progress-related information and then to provide customized responses to support users. In a simpler implementation, another study used shorter question formats for follow-up sessions [41].

The purposes of providing personalized content and conversations were to: (1) improve user engagement [35,37,38] and dialogue quality [42,43,54]; (2) provide timely feedback [39,40], adaptive user support [41], and adaptive training [36,38]; and, (3) support self-reflection [45,46].

Evaluation of Personalization

Only two studies directly assessed users' perceptions of personalization via custom-built questionnaires with questions on adaptive features [38] or via interview questions on tailored feedback [47]. One study employed a virtual coach to teach mindfulness and meditation [38]. The intervention group participants found the experience more rewarding, enjoyable, beneficial, and engaging than the control group participants. The coach's ability to provide customized feedback was the most successful feature, but this was only rated neutral to mildly positive (0.3 on a –2 to +2 Likert scale). Another study evaluated a relational agent to promote exercise and sun protection [47], with a total of 85% (29/34) of the study participants finding the tailored feedback helpful for achieving their behavior change goals [47]. The remaining studies did not directly evaluate the personalized features. Rather, they focused on evaluating factors that could be associated with personalization, such as user satisfaction, user engagement, and dialogue quality, or effects of personalization, such as improved skills, self-management, and awareness of the user's health status. One study conducted user interviews in which the users made positive remarks on personalization features [40].

Discussion

Principal Results

The use of CAs with unconstrained language input in health care is still limited, but there has been a notable increase in the number of studies in recent years. Almost half of the papers included in this study were published in the last two years. While most studies used quasi-experimental study designs, only two used randomized controlled trials [35,44]. Considering the recent emphasis on the role of replication in health informatics [55], the lack of technical details on conversational systems used in the studies is a major obstacle impeding replicability. In terms of personalization, our review found only 13 studies with personalization features. The studies provided various forms of personalized content, however, they were implemented without being supported by any prior evidence showing their effectiveness or any theoretical frameworks underpinning personalization [7,56,57]. Only three studies explicitly mentioned utilization of user profiles or user models to support personalized and adaptive features [42,43,46]. Similarly, only two studies directly assessed the personalization features [38,47]. The effects of the chosen personalization methods (either implicit, explicit, or a mix of the two) on user engagement and health outcomes received little attention.

While personalization of content to be delivered was common across all the studies, personalization of conversational style was implemented by only two studies [38,41]. The lack of conversational adaptation can be an impediment to improving usability and user experience, since different users may have different conversational preferences and needs that require different conversational strategies to be applied. Previous research has shown that adaptive conversational strategies can improve system performance, usability, and efficiency [58,59]. In this review, examples of conversational adaptation included using shorter questions for follow-up sessions [41] or using didactic, relational, or motivational conversational styles according to the user models [38]. Although such adaptive behaviors are useful steps towards accommodating the needs of various users, there are advanced implementations of conversational adaptation that can be applied to health care CAs, such as implicitly detecting users' level of expertise and thus adjusting the complexity of the terms used and the dialogue path to be taken [60], or configuring the level of system initiative and confirmation strategies when a user faces difficulties in performing a task [12].

Only two studies evaluated personalization as a distinct factor [38,47]. The direct assessment of personalization, involving how users perceive the extent of personalization, is an important element in the evaluation processes. When the effects of personalization were evaluated, it was not possible to determine whether any of the outcomes were attributed to the availability of personalization features or to other factors. Therefore, new conversational agent studies with carefully controlled conditions are needed to understand the relationships between personalization features and other evaluation factors such as user satisfaction and user engagement. To guide the direct assessment of personalization, a theoretical framework such as the one developed by Fan and Poole [7] may prove useful for systematically considering various dimensions of the personalization process.

The implications of different implementations of personalization were not addressed by any studies. For example, a recent research paper drew attention to the limitations of implicit and explicit personalization [61]: while implicit personalization with its often-imperceptible user models and hidden assumptions can result in biased decision-making [62], over-reliance on system suggestions [63], and filter bubbles [64], explicit personalization may involve very formulaic and superficial choices for users who may not be well-equipped to customize the presented choices in a satisfactory manner [65]. The study employed a reflective personalization approach, allowing users to reflect on their own goals and priorities when making or modifying choices [61]. This approach demonstrated an implementation of personalization that recognizes the complexities associated with human choices, preferences, and agency when using interactive technologies. Out of all the studies in our review, one study implemented a reflective approach to personalization by using adaptive mini-dialogues to support users' self-reflections on their goals [46]. These dialogues were successful in supporting discussions on

awareness, goal accomplishment, self-tracking data, and trends in behavior.

Using CAs with unconstrained natural language input can be risky [66]. Thus, it is important for such CAs to include patient harm considerations into their study protocols. None of the included studies reported any personalization-related harms., but there was a lack of attention to the safety implications of CAs, as evident in the absence of patient safety as an evaluation dimension. In addition to patient safety, future studies need to consider the effects of different personalization methods on patient privacy. In particular, implicit methods used for gathering user information need to be clearly communicated to the users, since such methods often run automatically in the background, not being noticed by users. To this end, the model of informed consent for information systems may prove useful for considering various factors involved in collecting personal information [67].

Overall, most of the reviewed papers did not focus explicitly on personalization. Little attention was generally paid to the complexities associated with implementing personalization features and measuring their effects.

Comparison with Prior Work

In line with our study, a recent scoping review of psychology-focused embodied conversational agents reported that only a few studies employed user models to personalize user-system interactions [68]. Another recent mapping study on health chatbots for behavior change found personalization to be one of the most appreciated technical enablers [69]. In terms of the implementation of personalization features, most of the studies in our review implemented personalization features without being informed by the advancements in other domains of personalization (eg, more automated personalization methods [12,60] or the implications of personalization on privacy, safety, and decision-making [61,62,70]).

Limitations

Our results are based on the presence of personalization features of health care CAs in the studies that do not necessarily have an explicit focus on personalization. Therefore, the results are limited by the extent to which the included studies reported on their personalization features. In addition, our review focused on CAs using unconstrained natural language input. Therefore, the results may not be extended to agents using constrained natural language input (eg, multiple-choice of utterance options). Since the conversational systems used in the reviewed studies involved multiple components, the reported outcomes were attributable to the systems rather than only the personalization features. Our paper recommended using a theoretical framework of personalization to support a more systematic treatment of personalization features. However, it may be possible to implement personalization features effectively with no theoretical support. Moreover, other theories not specific to personalization may prove useful for personalization purposes, such as the Theory of Planned Behavior [71]. Various contextual factors such as location and time may also be integrated with user models to support more adaptive information and services [72].

Future Research Directions

Future research can focus on incorporating a theoretical framework [7] and an evidence-based approach to implement personalization features in the domain of health care CAs. Another line of research could investigate the relationships between personalization features in conversational systems and health processes, and outcome measures such as treatment adherence or management of chronic health conditions. Future work can also focus on the use of the unique characteristics of the conversational medium for personalization purposes, such as capturing prosodic features in users' speech to automatically detect changes in mood or speech pathologies and thus provide adaptive information and services.

Conclusions

The use of personalization in health care CAs with unconstrained natural language interfaces has been limited and is not evidence based. While the CAs with personalization features were reported to improve user satisfaction, user engagement, and dialogue quality, little evaluation was performed to measure the extent of personalization and its role in improving health outcomes. Future research in health care CAs could evaluate the impact of personalization on health outcomes and its potential implications on privacy, safety, and decision-making.

Acknowledgments

This research was supported by the National Health and Medical Research Council (NHMRC) grant APP1134919 (Centre for Research Excellence in Digital Health). We would like to thank Catalin Tufanaru for his comments on the earlier drafts of this paper.

Authors' Contributions

This study was designed by ABK, JCQ, LL, DR, and EC. Search strategy was employed by ABK, LL, and HLT. Screening was performed by ABK, LL, and HLT. Data extraction was performed by ABK, SB, JCQ, LL, DR, HLT, and AB. First draft was written by ABK. Revisions and subsequent drafts were completed by ABK, SB, JCQ, LL, HLT, DR, AB, and EC.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The search strategy.
[PDF File (Adobe PDF File), 57 KB - jmir_v21i11e15360_app1.pdf]

Multimedia Appendix 2
The list of excluded articles.
[PDF File (Adobe PDF File), 158 KB - jmir_v21i11e15360_app2.pdf]

References

1. McTear M, Callejas Z, Griol D. The conversational interface: Talking to smart devices. Cham: Springer International Publishing; 2016.
2. Laranjo L, Dunn AG, Tong HL, Kocaballi AB, Chen J, Bashir R, et al. Conversational agents in healthcare: a systematic review. J Am Med Inform Assoc 2018 Sep 01;25(9):1248-1258 [FREE Full text] [doi: 10.1093/jamia/ocy072] [Medline: 30010941]
3. Pereira J, Díaz Ó. Using Health Chatbots for Behavior Change: A Mapping Study. J Med Syst 2019 Apr 04;43(5):135. [doi: 10.1007/s10916-019-1237-1] [Medline: 30949846]
4. Vaidyam AN, Wisniewski H, Halamka JD, Kashavan MS, Torous JB. Chatbots and Conversational Agents in Mental Health: A Review of the Psychiatric Landscape. Can J Psychiatry 2019 Jul;64(7):456-464. [doi: 10.1177/0706743719828977] [Medline: 30897957]
5. Montenegro J, da Costa C, da Rosa Righi R. Survey of conversational agents in health. Expert Systems with Applications 2019 Sep;129:56-67 [FREE Full text] [doi: 10.1016/j.eswa.2019.03.054]
6. Cambridge dictionaries online.: Cambridge University Press; 2015. URL: http://dictionary.cambridge.org [accessed 2019-09-16]
7. Fan H, Poole MS. What Is Personalization? Perspectives on the Design and Implementation of Personalization in Information Systems. Journal of Organizational Computing and Electronic Commerce 2006 Jan;16(3-4):179-202. [doi: 10.1080/10919392.2006.9681199]
8. Rich E. User modeling via stereotypes. In: Maybury MT, Wahlster W, editors. Readings in intelligent user interfaces. San Francisco, CA: Morgan Kaufmann Publishers Inc; 1998:329-342.
9. Pargellis A, Kuo H, Lee C. An automatic dialogue generation platform for personalized dialogue applications. Speech Communication 2004 Apr;42(3-4):329-351 [FREE Full text] [doi: 10.1016/j.specom.2003.10.003]
10. Kim Y, Bang J, Choi J, Ryu S, Koo S, Lee GG. Acquisition and Use of Long-Term Memory for Personalized Dialog Systems. : Springer International Publishing; 2015 Presented at: International Workshop on Multimodal Analyses Enabling Artificial Agents in Human-Machine Interaction; Sep 14 2014; Singapore p. 78-87 URL: https://doi.org/10.1007/978-3-319-15557-9_8 [doi: 10.1007/978-3-319-15557-9_8]
11. Thompson CA, Goker MH, Langley P. A Personalized System for Conversational Recommendations. Journal of Artificial Intelligence Research 2004;21:393-428. [doi: 10.1613/jair.1318]
12. Litman DJ, Pan S. Designing and evaluating an adaptive spoken dialogue system. User Modeling and User-Adapted Interaction 2002;2-3:111-137. [doi: 10.1023/A:1015036910358]
13. Aha DW, Breslow LA, Muñoz-Avila H. Conversational Case-Based Reasoning. J Applied Intelligence 2001;14(1):9-32. [doi: 10.1023/a:1008346807097]
14. Chen L, Pu P. Critiquing-based recommenders: survey and emerging trends. User Modeling and User-Adapted Interaction 2012;22(1-2):125-150. [doi: 10.1007/s11257-011-9108-6]
15. Levin E, Levin A. Evaluation of spoken dialogue technology for real-time health data collection. J Med Internet Res 2006 Dec 11;8(4):e30 [FREE Full text] [doi: 10.2196/jmir.8.4.e30] [Medline: 17213048]
16. Clark H, Wilkes-Gibbs D. Referring as a collaborative process. Cognition 1986 Feb;22(1):1-39 [FREE Full text] [doi: 10.1016/0010-0277(86)90010-7]
17. Nass C, Steuer J, Tauber ER. Computers are social actors. : ACM; 1994 Presented at: the SIGCHI Conference on Human Factors in Computing Systems; April 1994; Boston, Massachusetts, USA p. 72-78. [doi: 10.1145/259963.260288]
18. Brennan S. Lexical entrainment in spontaneous dialog. 1996 Presented at: Proceedings of ISSD 96; 1996; Philadelphia p. 41-44.
19. Carenini G, Moore JD. An empirical study of the influence of user tailoring on evaluative argument effectiveness. : Morgan Kaufmann Publishers Inc; 2001 Presented at: Proceedings of the 17th international joint conference on Artificial intelligence - Volume 2; 2001; Seattle, WA p. 1307-1312.
20. Demberg V, Winterboer A, Moore JD. A Strategy for Information Presentation in Spoken Dialog Systems. Computational Linguistics 2011 Sep;37(3):489-539. [doi: 10.1162/COLI_a_00064]
21. Walker M, Whittaker S, Stent A, Maloor P, Moore J, Johnston M, et al. Generation and evaluation of user tailored responses in multimodal dialogue. Cognitive Science 2004 Oct;28(5):811-840 [FREE Full text] [doi: 10.1016/j.cogsci.2004.06.002]

22. Pokorska-Bocci A, Stewart A, Sagoo GS, Hall A, Kroese M, Burton H. 'Personalized medicine': what's in a name? Per Med 2014 Mar;11(2):197-210. [doi: 10.2217/pme.13.107] [Medline: 29751382]
23. Noar SM, Grant Harrington N, Van Stee SK, Shemanski Aldrich R. Tailored Health Communication to Change Lifestyle Behaviors. American Journal of Lifestyle Medicine 2010 Nov 19;5(2):112-122. [doi: 10.1177/1559827610387255]
24. Hawkins RP, Kreuter M, Resnicow K, Fishbein M, Dijkstra A. Understanding tailoring in communicating about health. Health Educ Res 2008 Jun;23(3):454-466 [FREE Full text] [doi: 10.1093/her/cyn004] [Medline: 18349033]
25. Noar SM, Benac CN, Harris MS. Does tailoring matter? Meta-analytic review of tailored print health behavior change interventions. Psychol Bull 2007 Jul;133(4):673-693. [doi: 10.1037/0033-2909.133.4.673] [Medline: 17592961]
26. Kroeze W, Werkman A, Brug J. A systematic review of randomized trials on the effectiveness of computer-tailored education on physical activity and dietary behaviors. Ann Behav Med 2006 Jun;31(3):205-223. [doi: 10.1207/s15324796abm3103_2] [Medline: 16700634]
27. Cesuroglu T, Syurina E, Feron F, Krumeich A. Other side of the coin for personalised medicine and healthcare: content analysis of 'personalised' practices in the literature. BMJ Open 2016 Jul 13;6(7):e010243 [FREE Full text] [doi: 10.1136/bmjopen-2015-010243] [Medline: 27412099]
28. Coiera E. The true meaning of personalized medicine. Yearb Med Inform 2012;7:4-6. [Medline: 22890334]
29. Barnard KD, Lloyd CE, Dyson PA, Davies MJ, O'Neil S, Naresh K, et al. Kaleidoscope model of diabetes care: time for a rethink? Diabet Med 2014 May;31(5):522-530. [doi: 10.1111/dme.12400] [Medline: 24506524]
30. Revere D, Dunbar PJ. Review of computer-generated outpatient health behavior interventions: clinical encounters "in absentia". J Am Med Inform Assoc 2001;8(1):62-79 [FREE Full text] [doi: 10.1136/jamia.2001.0080062] [Medline: 11141513]
31. Bickmore T, Giorgino T. Health dialog systems for patients and consumers. J Biomed Inform 2006 Oct;39(5):556-571 [FREE Full text] [doi: 10.1016/j.jbi.2005.12.004] [Medline: 16464643]
32. Perez S. Techcrunch. 2019. Report: Voice assistants in use to triple to 8 billion by 2023 URL: https://techcrunch.com/2019/02/12/report-voice-assistants-in-use-to-triple-to-8-billion-by-2023/ [accessed 2019-09-16]
33. McTear MF. Spoken dialogue technology: enabling the conversational user interface. ACM Comput. Surv 2002;34(1):90-169. [doi: 10.1145/505282.505285]
34. Radziwill NM, Benton MC. Evaluating quality of chatbots and intelligent conversational agents. arXiv preprint 2017:1-21 [FREE Full text]
35. Fitzpatrick KK, Darcy A, Vierhile M. Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial. JMIR Ment Health 2017 Jun 06;4(2):e19 [FREE Full text] [doi: 10.2196/mental.7785] [Medline: 28588005]
36. Tanaka H, Negoro H, Iwasaka H, Nakamura S. Embodied conversational agents for multimodal automated social skills training in people with autism spectrum disorders. PLoS One 2017 Aug 10;12(8):e0182151 [FREE Full text] [doi: 10.1371/journal.pone.0182151] [Medline: 28796781]
37. Rhee H, Allen J, Mammen J, Swift M. Mobile phone-based asthma self-management aid for adolescents (mASMAA): a feasibility study. Patient Prefer Adherence 2014;8:63-72 [FREE Full text] [doi: 10.2147/PPA.S53504] [Medline: 24470755]
38. Hudlicka E. Virtual training and coaching of health behavior: example from mindfulness meditation training. Patient Educ Couns 2013 Aug;92(2):160-166 [FREE Full text] [doi: 10.1016/j.pec.2013.05.007] [Medline: 23809167]
39. Harper R, Nicholl P, McTear M, Wallace J, Black L, Kearney P. Automated Phone Capture of Diabetes Patients Readings with Consultant Monitoring via the Web. 2018 Presented at: the 15th Annual IEEE International Conference and Workshop on the Engineering of Computer Based Systems; March 31 - April 04, 2008; Belfast p. 219-226. [doi: 10.1109/ecbs.2008.31]
40. Black LA, McTear M, Black N, Harper R, Lemon M. Appraisal of a conversational artefact and its utility in remote patient monitoring. 2005 Presented at: 18th IEEE Symposium on Computer-Based Medical Systems; 23-24 June 2005; Dublin, Ireland. [doi: 10.1109/cbms.2005.33]
41. Levin E, Levin A. Evaluation of spoken dialogue technology for real-time health data collection. J Med Internet Res 2006 Dec 11;8(4):e30 [FREE Full text] [doi: 10.2196/jmir.8.4.e30] [Medline: 17213048]
42. Giorgino T, Azzini I, Rognoni C, Quaglini S, Stefanelli M, Gretter R, et al. Automated spoken dialogue system for hypertensive patient home management. Int J Med Inform 2005 Mar;74(2-4):159-167. [doi: 10.1016/j.ijmedinf.2004.04.026] [Medline: 15694621]
43. Azzini I, Falavigna D, Giorgino T, Gretter R, Quaglini S, Rognoni C, et al. Automated spoken dialog system for home care and data acquisition from chronic patients. Stud Health Technol Inform 2003;95:146-151. [Medline: 14663978]
44. Fulmer R, Joerin A, Gentile B, Lakerink L, Rauws M. Using Psychological Artificial Intelligence (Tess) to Relieve Symptoms of Depression and Anxiety: Randomized Controlled Trial. JMIR Ment Health 2018 Dec 13;5(4):e64 [FREE Full text] [doi: 10.2196/mental.9782] [Medline: 30545815]
45. Inkster B, Sarda S, Subramanian V. An Empathy-Driven, Conversational Artificial Intelligence Agent (Wysa) for Digital Mental Well-Being: Real-World Data Evaluation Mixed-Methods Study. JMIR Mhealth Uhealth 2018 Nov 23;6(11):e12106 [FREE Full text] [doi: 10.2196/12106] [Medline: 30470676]

46. Kocielnik R, Xiao L, Avrahami D, Hsieh G. Reflection Companion: A Conversational System for Engaging Users in Reflection on Physical Activity. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol 2018 Jul 05;2(2):1-26. [doi: 10.1145/3214273]
47. Sillice MA, Morokoff PJ, Ferszt G, Bickmore T, Bock BC, Lantini R, et al. Using Relational Agents to Promote Exercise and Sun Protection: Assessment of Participants' Experiences With Two Interventions. J Med Internet Res 2018 Feb 07;20(2):e48 [FREE Full text] [doi: 10.2196/jmir.7640] [Medline: 29415873]
48. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
49. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
50. Watson D, Clark LA, Tellegen A. Development and validation of brief measures of positive and negative affect: The PANAS scales. Journal of Personality and Social Psychology 1988;54(6):1063-1070. [doi: 10.1037//0022-3514.54.6.1063]
51. Hong H. Scale development for measuring health consciousness: Re-conceptualization. 2009 Presented at: 12th Annual International Public Relations Research Conference; 2009; Florida.
52. Walach H, Buchheld N, Buttenmüller V, Kleinknecht N, Schmidt S. Measuring mindfulness—the Freiburg Mindfulness Inventory (FMI). Personality and Individual Differences 2006 Jun;40(8):1543-1555 [FREE Full text] [doi: 10.1016/j.paid.2005.11.025]
53. Kember D, Leung DYP, Jones A, Loke AY, McKay J, Sinclair K, et al. Development of a Questionnaire to Measure the Level of Reflective Thinking. Assessment & Evaluation in Higher Education 2000 Dec;25(4):381-395. [doi: 10.1080/713611442]
54. Griol D, Carbó J, Molina JM. An Automatic Dialog Simulation Technique To Develop And Evaluate Interactive Conversational Agents. Applied Artificial Intelligence 2013 Oct 21;27(9):759-780. [doi: 10.1080/08839514.2013.835230]
55. Coiera E, Ammenwerth E, Georgiou A, Magrabi F. Does health informatics have a replication crisis? J Am Med Inform Assoc 2018 Aug 01;25(8):963-968 [FREE Full text] [doi: 10.1093/jamia/ocy028] [Medline: 29669066]
56. Blom J, Monk A. Theory of Personalization of Appearance: Why Users Personalize Their PCs and Mobile Phones. Human-Comp. Interaction 2003 Sep 1;18(3):193-228. [doi: 10.1207/s15327051hci1803_1]
57. Oulasvirta A, Blom J. Motivations in personalisation behaviour. Interacting with Computers 2008 Jan;20(1):1-16 [FREE Full text] [doi: 10.1016/j.intcom.2007.06.002]
58. Chu-Carroll J, Nickerson JS. Evaluating automatic dialogue strategy adaptation for a spoken dialogue system. : Association for Computational Linguistics; 2000 Presented at: North American Chapter of the Association for Computational Linguistics Conference; April 29 - May 04, 2000; Seattle, Washington p. 202-209.
59. Litman DJ, Pan S. Empirically Evaluating an Adaptable Spoken Dialogue System. In: Kay J, editor. UM99 User Modeling. CISM International Centre for Mechanical Sciences (Courses and Lectures)), vol 407. Vienna: Springer; 1999:55-64.
60. Maloor P, Chai J. Dynamic user level and utility measurement for adaptive dialog in a help-desk system. 2000 Presented at: the 1st SIGdial workshop on Discourse and dialogue-Volume 10; 7-8 October, 2000; Hong Kong, China p. 94-101. [doi: 10.3115/1117736.1117747]
61. Lee MK, Kim J, Forlizzi J, Kiesler S. Personalization revisited: a reflective approach helps people better personalize health services and motivates them to increase physical activity. 2015 Presented at: International Joint Conference on Pervasive and Ubiquitous Computing; September 07-11, 2015; Osaka, Japan p. 743-754. [doi: 10.1145/2750858.2807552]
62. Spalding E, Wilson A. Demystifying Reflection: A Study of Pedagogical Strategies That Encourage Reflective Journal Writing. Teachers College Rec 2002 Oct;104(7):1393-1421. [doi: 10.1111/1467-9620.00208]
63. Carr N. The glass cage: Where automation is taking us. London: Random House; 2015.
64. Pariser E. The filter bubble: how the new personalized Web is changing what we read and how we think. Choice Reviews Online 2012 Oct 01;50(02):50-0926-50-0926. [doi: 10.5860/choice.50-0926]
65. Riquelme H. Do consumers know what they want? Journal of Consumer Marketing 2001 Sep;18(5):437-448. [doi: 10.1108/07363760110398772]
66. Bickmore TW, Trinh H, Olafsson S, O'Leary TK, Asadi R, Rickles NM, et al. Patient and Consumer Safety Risks When Using Conversational Assistants for Medical Information: An Observational Study of Siri, Alexa, and Google Assistant. J Med Internet Res 2018 Sep 04;20(9):e11510 [FREE Full text] [doi: 10.2196/11510] [Medline: 30181110]
67. Friedman B, Lin P, Miller JK. Informed consent by design. In: Cranor LF, Garfinkel S, editors. Security and Usability: Designing Secure Systems that People Can Use. Farnham, UK: O'Reilly Media; 2005:495-522.
68. Provoost S, Lau HM, Ruwaard J, Riper H. Embodied Conversational Agents in Clinical Psychology: A Scoping Review. J Med Internet Res 2017 May 09;19(5):e151 [FREE Full text] [doi: 10.2196/jmir.6553] [Medline: 28487267]
69. Pereira J, Díaz Ó. Using Health Chatbots for Behavior Change: A Mapping Study. J Med Syst 2019 Apr 04;43(5):135. [doi: 10.1007/s10916-019-1237-1] [Medline: 30949846]
70. Awad, Krishnan. The Personalization Privacy Paradox: An Empirical Evaluation of Information Transparency and the Willingness to Be Profiled Online for Personalization. MIS Quarterly 2006;30(1):13. [doi: 10.2307/25148715]
71. Ajzen I. The theory of planned behavior. Organizational Behavior and Human Decision Processes 1991 Dec;50(2):179-211 [FREE Full text] [doi: 10.1016/0749-5978(91)90020-t]

72. Jameson A. Modelling both the Context and the User. Personal and Ubiquitous Computing 2001 Feb 28;5(1):29-33. [doi: 10.1007/s007790170025]

Abbreviations

CA: Conversational Agent
FMI: Freiburg Mindfulness Inventory
GAD: Generalized Anxiety Disorder
NR: Not reported
PANAS: Positive and Negative Affect Scale
PHQ: Patient Health Questionnaire
RQ: Reflection Questionnaire

Edited by G Eysenbach; submitted 04.07.19; peer-reviewed by H Tanaka, R Kocielnik, K Fauvelle; comments to author 20.08.19; accepted 03.09.19; published 07.11.19.

Please cite as:
Kocaballi AB, Berkovsky S, Quiroz JC, Laranjo L, Tong HL, Rezazadegan D, Briatore A, Coiera E
The Personalization of Conversational Agents in Health Care: Systematic Review
J Med Internet Res 2019;21(11):e15360
URL: https://www.jmir.org/2019/11/e15360
doi:10.2196/15360
PMID:31697237

Original Paper

Conversational Interfaces for Health: Bibliometric Analysis of Grants, Publications, and Patents

Zhaopeng Xing[1], MPS; Fei Yu[1,2], PhD, MPS; Jian Du[3], PhD; Jennifer S Walker[2], MSIS; Claire B Paulson[1], BA; Nandita S Mani[2], PhD; Lixin Song[4,5], PhD

[1]Carolina Health Informatics Program, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

[2]Health Science Library, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

[3]National Institute of Health Data Science, Peking University, Beijing, China

[4]School of Nursing, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

[5]Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

Corresponding Author:
Zhaopeng Xing, MPS
Carolina Health Informatics Program
University of North Carolina at Chapel Hill
Health Science Library, 335 S Columbia St
Chapel Hill, NC
United States
Phone: 1 919 260 5844
Email: zhaopeng@live.unc.edu

Abstract

Background: Conversational interfaces (CIs) in different modalities have been developed for health purposes, such as health behavioral intervention, patient self-management, and clinical decision support. Despite growing research evidence supporting CIs' potential, CI-related research is still in its infancy. There is a lack of systematic investigation that goes beyond publication review and presents the state of the art from perspectives of funding agencies, academia, and industry by incorporating CI-related public funding and patent activities.

Objective: This study aimed to use data systematically extracted from multiple sources (ie, grant, publication, and patent databases) to investigate the development, research, and fund application of health-related CIs and associated stakeholders (ie, countries, organizations, and collaborators).

Methods: A multifaceted search query was executed to retrieve records from 9 databases. Bibliometric analysis, social network analysis, and term co-occurrence analysis were conducted on the screened records.

Results: This review included 42 funded projects, 428 research publications, and 162 patents. The total dollar amount of grants awarded was US $30,297,932, of which US $13,513,473 was awarded by US funding agencies and US $16,784,459 was funded by the Europe Commission. The top 3 funding agencies in the United States were the National Science Foundation, National Institutes of Health, and Agency for Healthcare Research and Quality. Boston Medical Center was awarded the largest combined grant size (US $2,246,437) for 4 projects. The authors of the publications were from 58 countries and 566 organizations; the top 3 most productive organizations were Northeastern University (United States), Universiti Teknologi MARA (Malaysia), and the French National Center for Scientific Research (CNRS; France). US researchers produced 114 publications. Although 82.0% (464/566) of the organizations engaged in interorganizational collaboration, 2 organizational research-collaboration clusters were observed with Northeastern University and CNRS as the central nodes. About 112 organizations from the United States and China filed 87.7% patents. IBM filed most patents (N=17). Only 5 patents were co-owned by different organizations, and there was no across-country collaboration on patenting activity. The terms *patient*, *child*, *elderly*, and *robot* were frequently discussed in the 3 record types. The terms related to mental and chronic issues were discussed mainly in grants and publications. The terms regarding multimodal interactions were widely mentioned as users' communication modes with CIs in the identified records.

Conclusions: Our findings provided an overview of the countries, organizations, and topic terms in funded projects, as well as the authorship, collaboration, content, and related information of research publications and patents. There is a lack of broad cross-sector partnerships among grant agencies, academia, and industry, particularly in the United States. Our results suggest a need to improve collaboration among public and private sectors and health care organizations in research and patent activities.

(J Med Internet Res 2019;21(11):e14672) doi:10.2196/14672

KEYWORDS

conversational interfaces; conversational agents; chatbots; artifical intelligence; healthcare; bibliometrics; social network; grants; publications; patents

Introduction

The emergence of conversational interfaces (CIs) enables users to talk to a machine [1]. Using conventional pattern match or natural language processing, CIs simulate human conversation through various interaction modalities [2-4]. For example, text-based CIs, also known as chatbots, are commonly presented on messaging platforms where users can converse with bots using textual input (eg, Facebook Messenger and Slack). Voice-based CIs incorporate a speech channel into the interface and are preferred over traditional graphical interfaces (eg, keyboard and screen) in visual or hands-off tasks [5]. Recent examples of voice-based CIs include intelligent personal assistants (eg, Apple Siri, Amazon Alexa, and Microsoft Cortana) and CIs with multimodalities such as embodied conversational agents (ECAs). Multimodality can be designed to engage verbal and nonverbal interactions (eg, gaze movement, facial expression, and gesture) between the CI system and users [6,7]. This mode can also be implemented using an embodied character and menu-based dialogue module to simulate a dialogue flow [8].

In health care, CIs have been adopted to provide complementary therapy and health behavioral interventions [9-11], assist patient self-management [12-15], and support clinical decision making [16-18]. With the growth of CIs for health, there have been increasing efforts in appraising research in CI design and applications. Laranjo et al [2] identified 17 studies that have investigated 14 unique CI applications for general health-related purposes. These studies reported that CIs have produced positive outcomes related to patient engagement and adherence and decreased self-reported symptoms. A review examined 40 studies regarding CI technologies in hospital settings and proposed a taxonomy that involved interaction context, dialogue types, and architecture attributes. Researchers reported that CIs were primarily designed for physician education and patient counseling purposes [3]. A review of 8 studies that applied CIs to treat mental illness [4] reported positive outcomes and user satisfaction yet inconsistent evaluation of CI technologies.

Despite the rapidly growing research evidence supporting the potentials of CI applications in health care, there is a lack of systematic investigation that goes beyond the published research and incorporates grant and patent activities. This will demonstrate the state of the art from the perspectives of funding agencies, academia, and industry. Research grants reflect funding agencies' attention to and support for emerging research domains [19,20]. Publications and patents have been widely used to map the emergence of technologies and reveal research and development (R&D) activities [21,22]. A broad scope review using data from multiple sources, such as scientific publications, funding instruments, and patents, will thus provide an overview of the domain landscape and inform stakeholders [23,24] for better CI utilization and research. In addition, a recent report on artificial intelligence (AI) by World Intelligence Property Organization (WIPO) [25] identified the untapped opportunities that AI technologies have brought to health care and called for a broad collaboration and/or coordination among funding agencies, policymakers, researchers, and entrepreneurs. As one of the most promising health care applications of AI, CIs should be given sufficient attention. This study aimed to use data systematically extracted from multiple sources (ie, grant, publication, and patent databases) to investigate the development and research of health CIs. We examined the following 5 research questions:

- RQ1: Basic statistics of the collected records: how many grants, publications, and patents exist for CIs used for health purposes?
- RQ2: Analysis of grants: which funding agencies have granted the largest amount of funds in CIs and which organizations have received funds in CIs?
- RQ3: Analysis of research publications: who are the top contributors to research publications, who are their main collaborators, and how are they distributed geographically?
- RQ4: Analysis of patents: where were the patents filed (ie, country) and who were the most active patent assignees (ie, organizations)?
- RQ5: Analysis of topic terms: what terms were frequently addressed in grants, publications, and patents and what potential gaps can be identified for future research?

Methods

Data Collection

To identify grants, publications, and patents, we created a search strategy adapted from previously published works that include CI-related terms [1,2,21,26], such as *spoken dialogue system*, *conversational agent*, *chatbot*, *social robot*, *virtual agent*, *and question-answer system.* The health-related terms covered both generic health term variations and key stakeholders, such as *health*, *healthcare*, *medicine*, clinic, *physician*, *patient*, and *caregiver*. In addition, we used a snowball process to collect synonyms, term variations, and relevant terms to build a search term thesaurus for this study [27]. Search queries were modified for different databases (see Multimedia Appendix 1).

We systematically searched 9 databases to identify granted projects, research publications, and patents from January 2008 to December 2018 that were for health purposes (Table 1). Initial searches were conducted in August 2018 and updated in January 2019. The databases for grant search included Federal RePORTER and Community Research and Development Information Service (CORDIS). Federal RePORTER is a publicly available database of scientific funding projects by federal agencies, including National Institutes of Health (NIH), National Science Foundation (NSF), the Agency for Healthcare Research and Quality (AHRQ), Department of Agriculture,

National Aeronautics and Space Administration, Department of Defense, Environmental Protection Agency, and Department of Education. For a broader coverage of funding records, we also searched the Europe Commission (EC) by using CORDIS, a primary information source of funded projects in the European Union countries.

The publication search was conducted in 6 bibliographic databases that index literature in technology, biomedical, and health sciences (Table 1). Publications were limited to conference papers, journal articles, and book sections and excluded meeting abstracts, editorials, letters, or lecture notes.

We retrieved the filed and granted patents from Derwent Innovation Index (DII). Sponsored by Clarivate Analytics, DII is a widely used database in patent bibliometric analysis and has a broad coverage of over 37 million inventions and 70 million patents from 52 authorities in the world [24,28].

After searches were completed, the retrieved records from the 9 selected databases were aggregated and deduped in EndNote (Clarivate Analytics). Particularly, if a record had a duplication, we only kept the one which had complete data fields. We also cross-checked other data resources for the missing parts of the records.

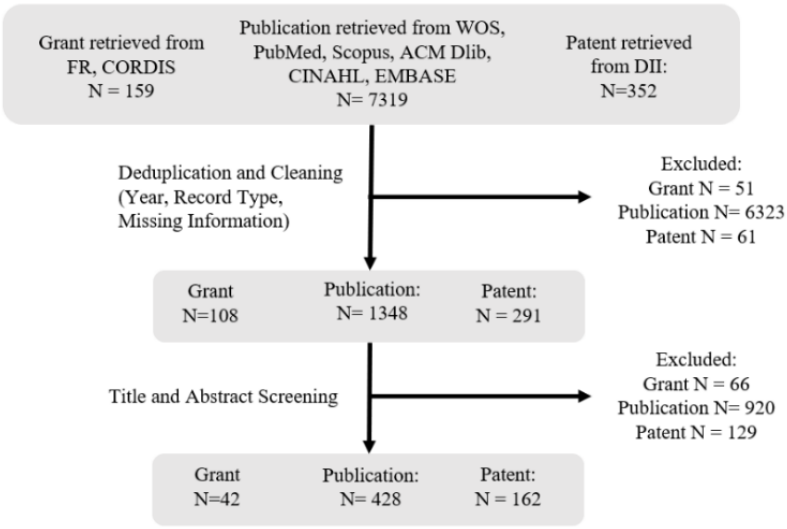
Table 1. Databases of grants, publications, and patents.

Data type	Database searched
Grants	Federal RePORTER and Community Research and Development Information Service
Publications	Web of Science, PubMed, Scopus, EMBASE, Association for Computing Machinery Digital Library, and Cumulative Index of Nursing and Allied Health
Patents	Derwent Innovation Index

Data Screening

We included grants, scientific publications, and patents that were relevant to the design, development, improvement, deployment, or evaluation of CIs for health purposes. A record was excluded if (1) it was awarded, published, or filed before 2008; (2) on cross-check, it missed any of the following fields: organization recipients of a grant, grant agency, grant size, or budget year; title, abstract, or author affiliation; and country, patent assignee, or priority date; and (3) it was irrelevant to CI technologies for health purposes. For example, we included the records that reported systems, programs, and interfaces, which enabled users to issue command requests or engage in dialogue/chats in different input modalities (eg, text based, spoken, menu selection, or multimodal). We excluded the records that did not involve or specify the abovementioned modes, such as pet robot, nonverbal virtual robots, and the Internet of Things. We also excluded the records that examined CIs for nonhealth purposes. For patents, we included the records that claimed CI inventions that can be used for health purposes. After deduping and cleaning the data, ZX manually screened the titles and abstracts of the records, and the records with uncertainty were discussed at team meetings and resolved by co-authors (Figure 1).

Figure 1. Data collection and screening.



Data Analysis

We employed bibliometric analysis, social network analysis (SNA), and term co-occurrence analysis techniques to analyze the abstracted data (Table 2). Bibliometric analysis is the quantitative analysis of scientific publications [29]. It has been widely used for measuring research and patenting performance [21,30] and recently extended to measuring research funding dynamics [23,24]. SNA was originally developed for social structure study. Using bibliographic data, SNA expands the scope of bibliometrics by revealing the co-authorship among different research units [31,32]. Built on network theory and bibliometrics, the term co-occurrence analysis represents the knowledge components of a document by key terms and their co-occurrence relationship [33]. Researchers have used the term

co-occurrence approach to explore potential research topics in scientific publications [34,35].

Table 2. Extracted data fields, analysis methods, and metrics.

Data sample	Data fields	Analysis methods	Metrics
Grant	Title, abstract, grant agency, granted organization, grant size, and grant start year	Bibliometric analysis, social network analysis, and term co-occurrence analysis	Number of funded projects, grant size, project counts, project duration, funding and recipient agency and organizations, and topic co-occurrence
Publication	Title, abstract, affiliation, and publication year	Bibliometric analysis, social network analysis, and term co-occurrence analysis	Number of publications, organizational network, and topic co-occurrence
Patent	Title, abstract, patent assignee, and the priority date[a]	Bibliometric analysis, social network analysis, and term co-occurrence analysis	Number of patents, organizational collaboration network, and topic co-occurrence

[a]We used the priority date as the time stamp for patent data, which is the date when the first patent application in a patent family was filed at a patent office (priority application) and which is often used to establish the priority of an invention [36].

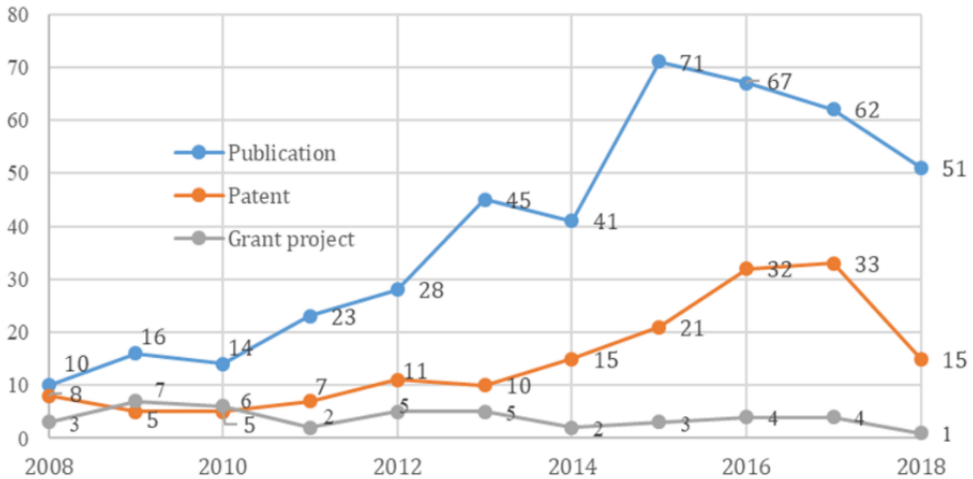
We used both VOSviewer (version 1.6.9) and Tableau for data analysis and visualization. VOSviewer is a bibliometric network analysis software to conduct SNA for collaboration networks and term co-occurrence analysis for topic analysis [37,38]. For SNA, each organization was represented as a node, and each collaboration between 2 nodes was represented by a link. For the term co-occurrence analysis, key terms were extracted from the title and abstract of the records. We used a density map to visualize the disclosed key terms. The importance of a topic was measured by frequency of terms (ie, counts of a term) and its co-occurrence with other terms. The synonyms of extracted key terms were consolidated for visualization. For example, the extracted terms *child*, *children*, and *kid* were merged and standardized as *child*.

Results

The Number of Grants, Research Publications, and Patents

After data screening, we obtained a dataset with 42 grants, 428 research publications, and 162 patents related to CIs for health purposes. As displayed in Figure 2, the number of grants has remained steady from 2008 to 2018, with an average of 3.8 projects per year. The number of publications increased since 2010, with an average growth rate of 22.1% per year, which then reached a peak in 2015. The number of patents also grew from 2010 to 2017, with an annual growth rate of 22%. The US patent applications filed on and after 2018 are not publicly available until 18 months after the application's earliest filing date [36], causing the drop-in number of patent records in 2018.

Figure 2. Number of grants, research publications, and patents between 2008 and 2018.



Analysis of Grants

Among the 42 funded research projects focusing on CI technologies for health purposes, the total dollar amount of grants awarded was US $30,297,932, of which US $13,513,473 was awarded by funding agencies in the United States, and US $16,784,459 was awarded by the EC (Table 3 and Figure 3).

The top 3 funding agencies in the United States were NSF, NIH, and AHRQ, which overall funded 45.2% of the included projects. Among the 6 EC-funded projects identified in this study, the EC grants were allocated to 10 countries, that is, the United Kingdom, Spain, France, Italy, Germany, Norway, Belgium, Israel, Latvia, and Romania (Table 4), for either individual or collaborative projects.

Table 3. Grant count, size, and average duration of projects.

Agency	Project count	Percentage of total count	Grant size	Percentage of total grants	Months of project, mean (SD)
National Science Foundation (the United States)	19	45.2	US $8,686,669	28.67	33.9 (15.9)
National Institutes of Health (the United States)	16	38.1	US $4,645,612	15.33	34.9 (17.4)
Agency for Healthcare Research and Quality (the United States)	1	2.4	US $181,192	0.60	24.0 (N/A[a])
Europe Commission (European Union)	6	14.3	US $16,784,459	55.40	33.0 (6.7)

[a]Not applicable.

Figure 3. Grant size (USD) and project duration (months) from 2008 to 2018. AHRQ: Agency for Healthcare Research and Quality; EC: Europe Commission; NIH: National Institutes of Health; NSF: National Science Foundation.

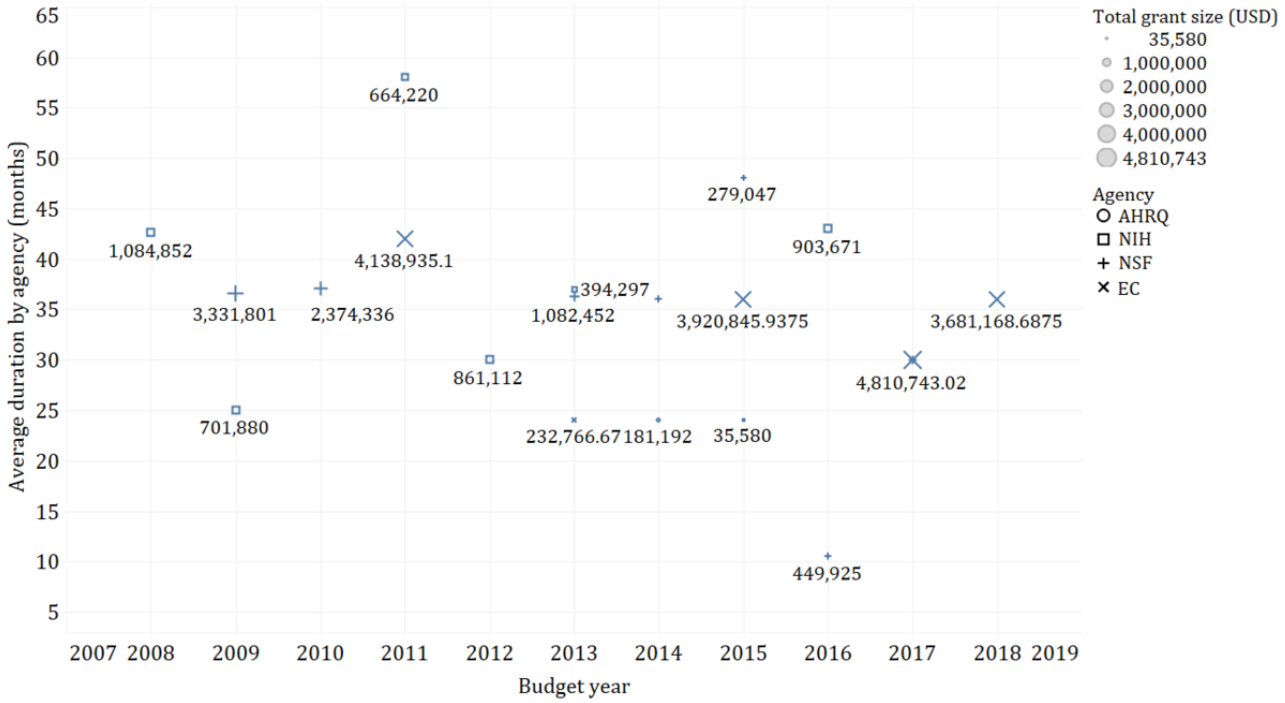


Table 4. The grant size and number of grant recipients by country.

Country	Organization recipients count	Total grant size
United States	27	US $13,513,473
United Kingdom	9	US $3,914,721
Spain	8	US $3,048,271
France	5	US $2,350,823
Italy	5	US $2,567,397
Germany	4	US $2,120,652
Norway	2	US $1,392,323
Belgium	1	US $330,315
Israel	1	US $218,738
Latvia	1	US $639,540
Romania	1	US $201,680

The 42 funded projects involve a total of 64 granted organizations (Figure 4). Among the 27 US grantee organizations, 26 were research institutes. Boston Medical Center was awarded the largest grant size with 3 projects funded by NIH with a total amount of US $2,065,245 and 1 project funded by AHRQ with the amount of US $181,192. These 4 projects were granted for designing and evaluating CIs to improve patients' engagement [39], for palliative care of patients

with advanced illness, for reducing cardiopulmonary rehospitalization [40], and for treating comorbid depression (no outcomes reported). In Europe, the EC funded 18 private companies and 19 academic institutes. The University of Edinburg received the largest grant for a project that applied CI technology to support the treatment of major depression (ie, US $1,758,786) [41].

The project duration varied from 3 to 69 months. For example, Auburn University and the University of South Florida were granted 3-month funding from NSF and NIH, respectively. The University of Delaware received funding that lasted for 69 months.

Figure 4. Grant agencies and granted organizations network. AHRQ: Agency for Healthcare Research and Quality; EC: Europe Commission; NIH: National Institutes of Health; NSF: National Science Foundation.

Analysis of Research Publications

This study identified a total of 428 publications authored by researchers from 58 countries. As shown in Table 5, the researchers from the United States published 114 manuscripts on CI technologies for health purposes. The other countries that have published the most about this line of work included Japan ($n_{publication}$=34), France ($n_{publication}$ =33), China ($n_{publication}$=28), the United Kingdom ($n_{publication}$=24), Italy ($n_{publication}$=21), Malaysia ($n_{publication}$=20), Germany ($n_{publication}$=17), and the Netherlands ($n_{publication}$=16). Overall, researchers in European countries produced about 31.1% of the total publications compared with 19.2% by researchers in Asian countries.

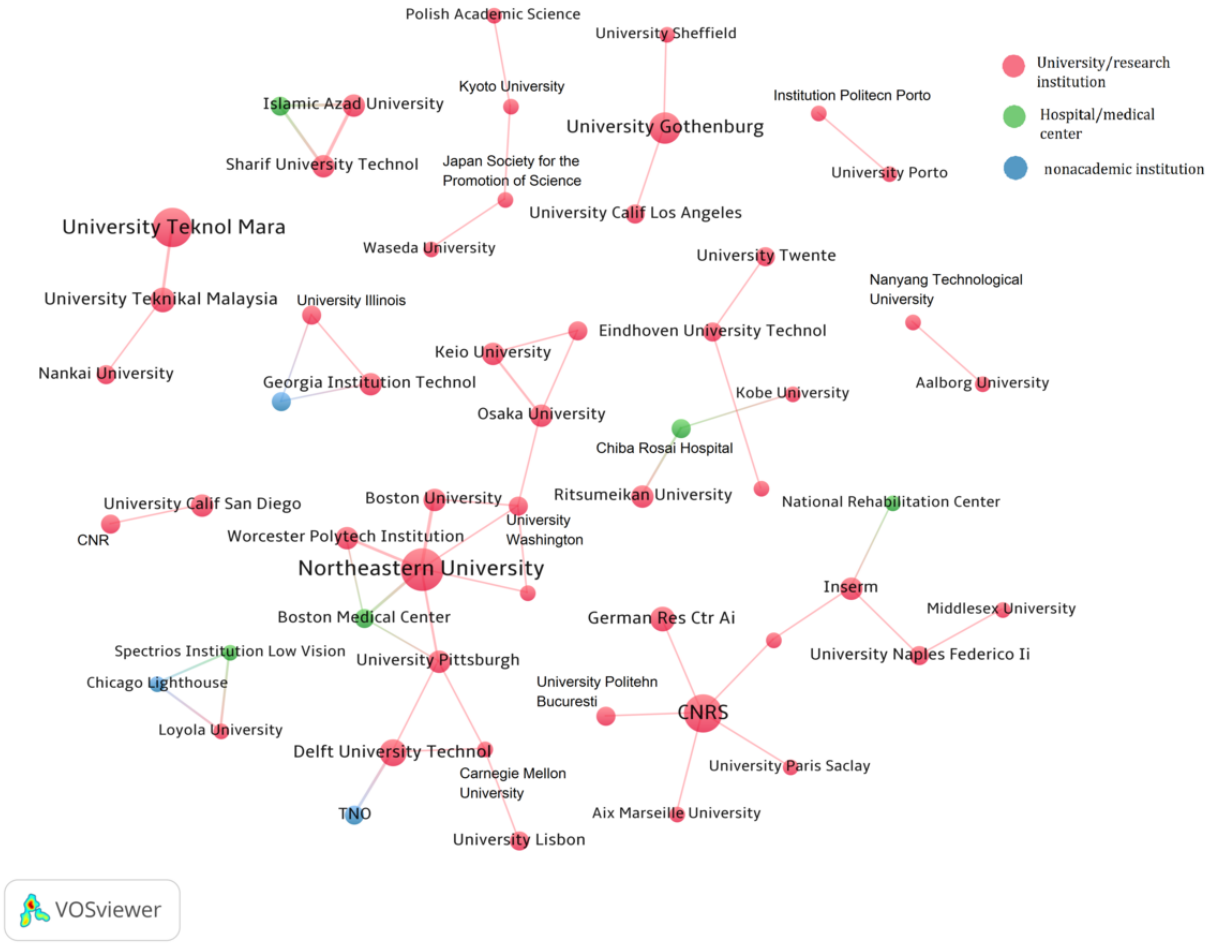
Among the 566 organizations with which the authors were affiliated, 95 organizations (16.8%) contributed at least two publications, including 85 academic institutions, 6 hospitals, and 4 companies or private organizations. As shown in Table 5, the top 3 most productive institutes were Northeastern University, the Universiti Teknologi MARA System in Malaysia, and the French National Center for Scientific Research (also known as and shown as CNRS).

Regarding organizational collaboration, 246 of 428 publications (57.5%) were co-authored by researchers from at least two organizations. Overall, 463 organizations participated in interorganizational collaborations, whereas 303 organizations had more than 1 external collaborator. The top 5 institutes were CNRS (20 collaborators), Inserm (14 collaborators), Osaka University (13 collaborators), Northeastern University (11 collaborators), and the University of Naples Federico II (11 collaborators; Figure 5). A total of 2 collaboration clusters were identified from the organizational collaboration network with Northeastern University and CNRS as the central nodes.

Table 5. The number of publications by country and organization (Top 10), N=248.

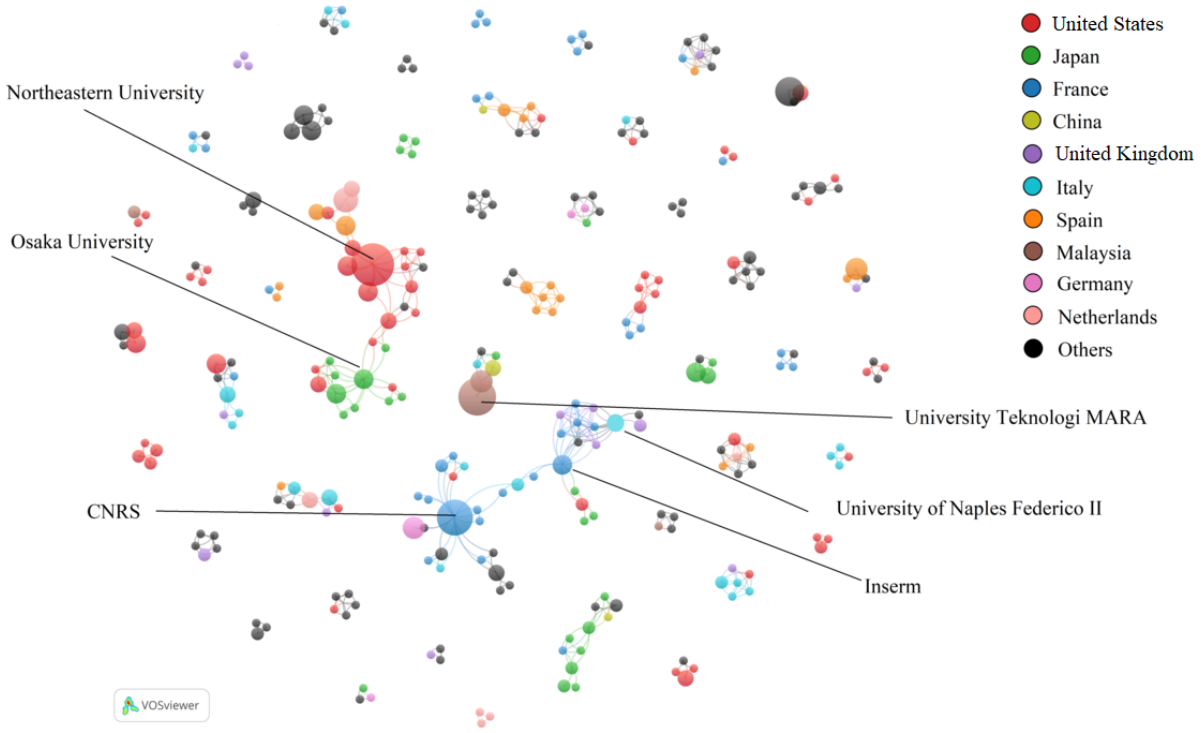
Rank	Publications by country		Publication by organization	
	Country	Publication count, n (%)	Organization	Publication count, n (%)
1	United States	114 (26.6)	Northeastern University (US)	15 (3.5)
2	Japan	34 (7.9)	University Teknologi MARA (Malaysia)	13 (3.0)
3	France	33 (7.7)	CNRS (France)	12 (2.8)
4	China	28 (6.5)	University of South California (US)	9 (2.1)
5	United Kingdom	24 (5.1)	University of Gothenburg (Sweden)	8 (1.9)
6	Italy	21 (4.9)	Delft University of Technology (Netherlands)	6 (1.4)
7	Spain	21 (4.9)	Istanbul Technical University (Turkey)	6 (1.4)
8	Malaysia	20 (4.7)	University Carlos III de Madrid (Spain)	5 (1.2)
9	Germany	17 (4.0)	Universiti Teknikal Malaysia Melaka (Malaysia)	5 (1.2)
10	Nether-lands	16 (3.7)	German Research Center for Artificial Intelligence (Germany)	5 (1.2)

Figure 5. Publication organizational co-authorship network at country level. Each node represents an organization, and the size of the node represents publication count. CNRS: National Center for Scientific Research.



In addition to academic institutions, 5 hospitals/medical centers (ie, Chiba Rosal Hospital, Mahak Hospital, Boston Medical Center, Spectrios Institute for Low Vision, and National Rehabilitation Center for Persons with Disabilities in Japan), and 3 nonacademic institutions (ie, TNO, IBM, and Chicago Lighthouse) were involved in the collaboration network. The only collaboration across 3 types of organizations was among Loyola University, Chicago Lighthouse (a nonprofit social service organization), and the Spectrios Institute for Low Vision (Figure 6).

Figure 6. Publication co-authorship network at the organizational level. The organizations shown are those with more than 1 publication.



Analysis of Patents

Among the 162 patents reviewed in this study, the United States and China filed the majority of the patent inventions, 78 and 69 respectively, and they are followed by Japan (n_{patent} =7) and South Korea (n_{patent} =6). Regarding the patent assignees, 112 organizations were identified, including 98 private companies and 14 research institutes. Among research institutes, 9 were based in China, 2 in Japan, 2 in the United States, and 1 in South Korea. IBM filed more patents (n_{patent} =17) than all the other organizations or companies. Organizations that filed more than 2 patents were Vocollect Healthcare (n_{patent} =6), Google (n_{patent} =4), Next It Corp (n_{patent} =3) and Samsung (n_{patent} =3).

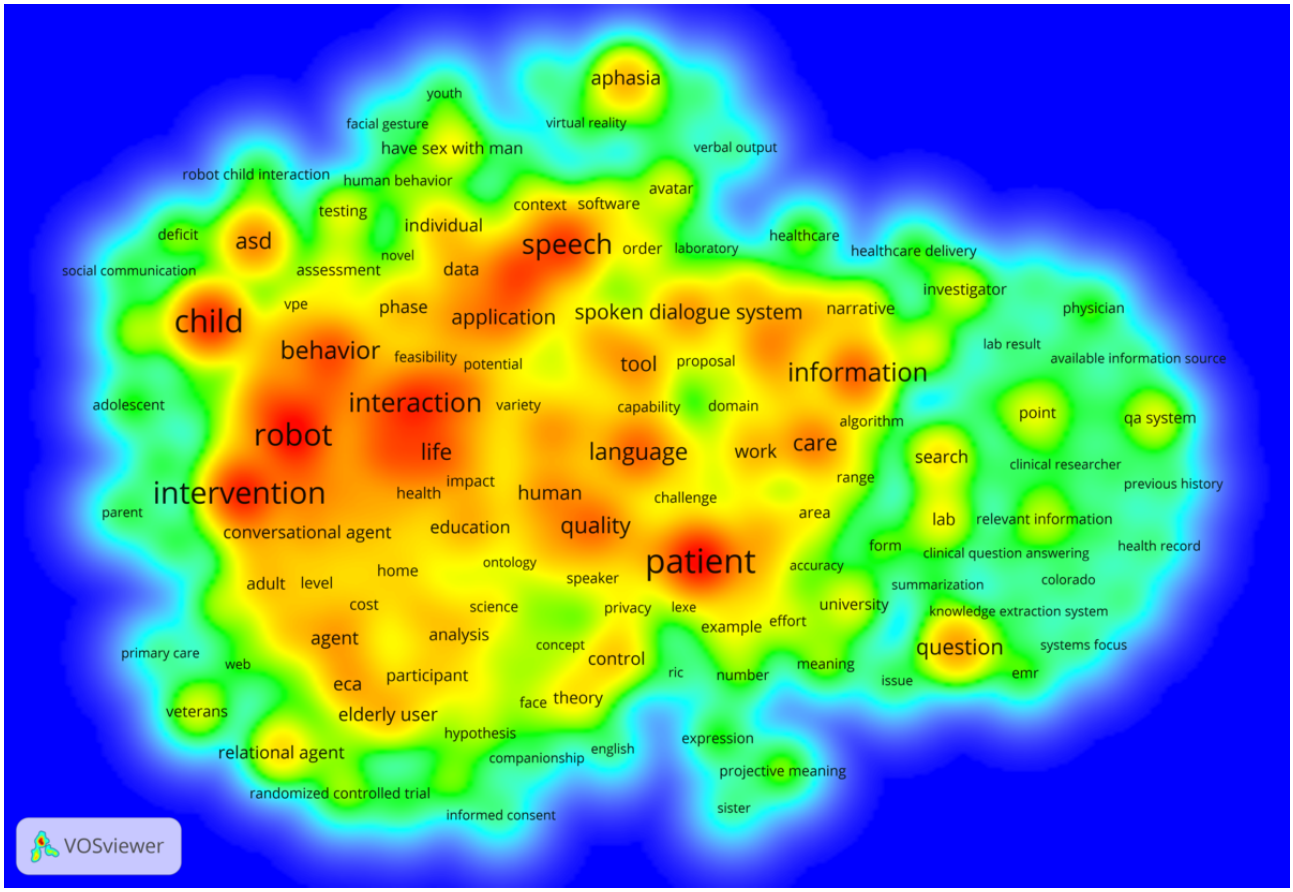
With regard to the collaboration among patent assignees, we only identified 5 collaboration patents that were co-assigned to 5 pairs of organizations in the United States, Japan, and China. They included (1) Puretech Management Inc and Bose Corp (United States), (2) Next It Corp and Verint Americas Inc (United States), (3) University of Kyoto and Toyota (Japan), (4) Jiangsu province Hospital and Nanjing Zhongyue Information Technology Company (China), and (5) University of South China Technology and Guanzhou Lvsong Biotechnology Company (China). Furthermore, this study did not find any cross-country patent collaborations among identified assignees.

Analysis of Topic Terms

Grants

The most commonly occurring terms (Figure 7) included *patient*, *child*, *intervention*, *robot*, and speech. The grants primarily focused on the *patient*, and 7 projects targeted the *child* or *elderly* populations. In addition, the term *asd* (ie, autism spectrum disorder [ASD]) was in very close proximity with *child* and occurred in 5 funded projects. For example, an NIH-sponsored project at the University of Connecticut and the University of Delaware in 2009 used social robots to support children with ASD. The term *elderly* was also addressed in 7 grants. For example, Auburn University and Clemson University proposed a project, funded by NSF in 2009, aimed to use CI to improve older adults' quality of life. Furthermore, *speech* seems to be a preferred interaction modality in the proposals. For example, the University of Texas Dallas was funded by NSF to create a speech-enabled CI, which aimed to help individuals with hearing impairment and autistic children improve communication skills.

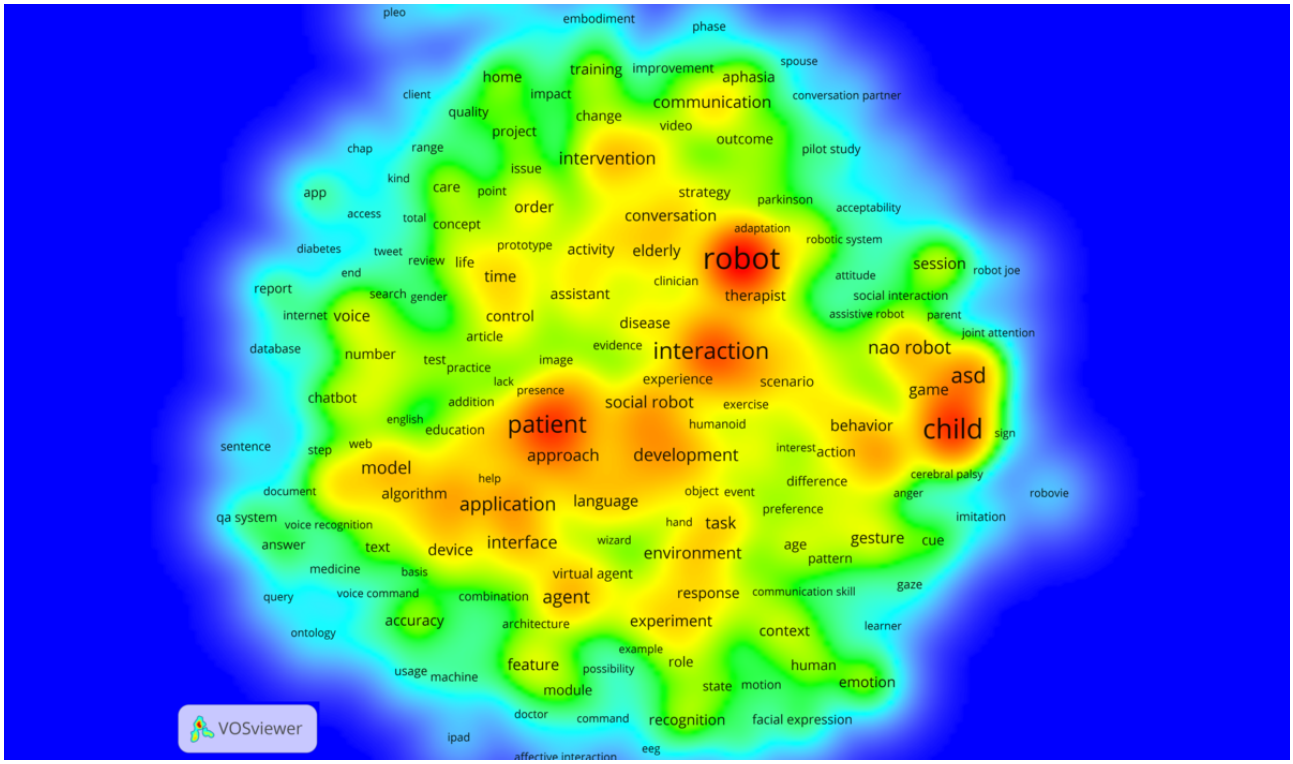
Figure 7. Heatmap of the topic terms that occurred in more than 4 granted projects.



Publications

The terms with top occurrence in publications were *robot*, *patient*, and *child* (Figure 8). In addition, *asd* appeared together with *child* in 46 publications [42,43]. Other disease-related terms discussed in the publications included *aphasia* (3.7%, 16/428) [44,45], *cerebral palsy* (2.8%, 12/428) [46,47], *stroke* (2.6%, 11/428) [48,49], *cancer* (1.4%, 6/428) [50,51]. The key term *elderly* was addressed in 42 publications [52,53]. The term *multimodal* occurred in 39 publications [54,55], and specific terms associated with multimodal interaction were *gesture* (10.3%, 44/428), *voice* (10.0%, 43/428), and *facial expression* (5.1%, 22/428). Furthermore, *NAO robot*, a humanoid robot developed by SoftBank Robotics, was investigated in 87 publications (20.3%) [42,43].

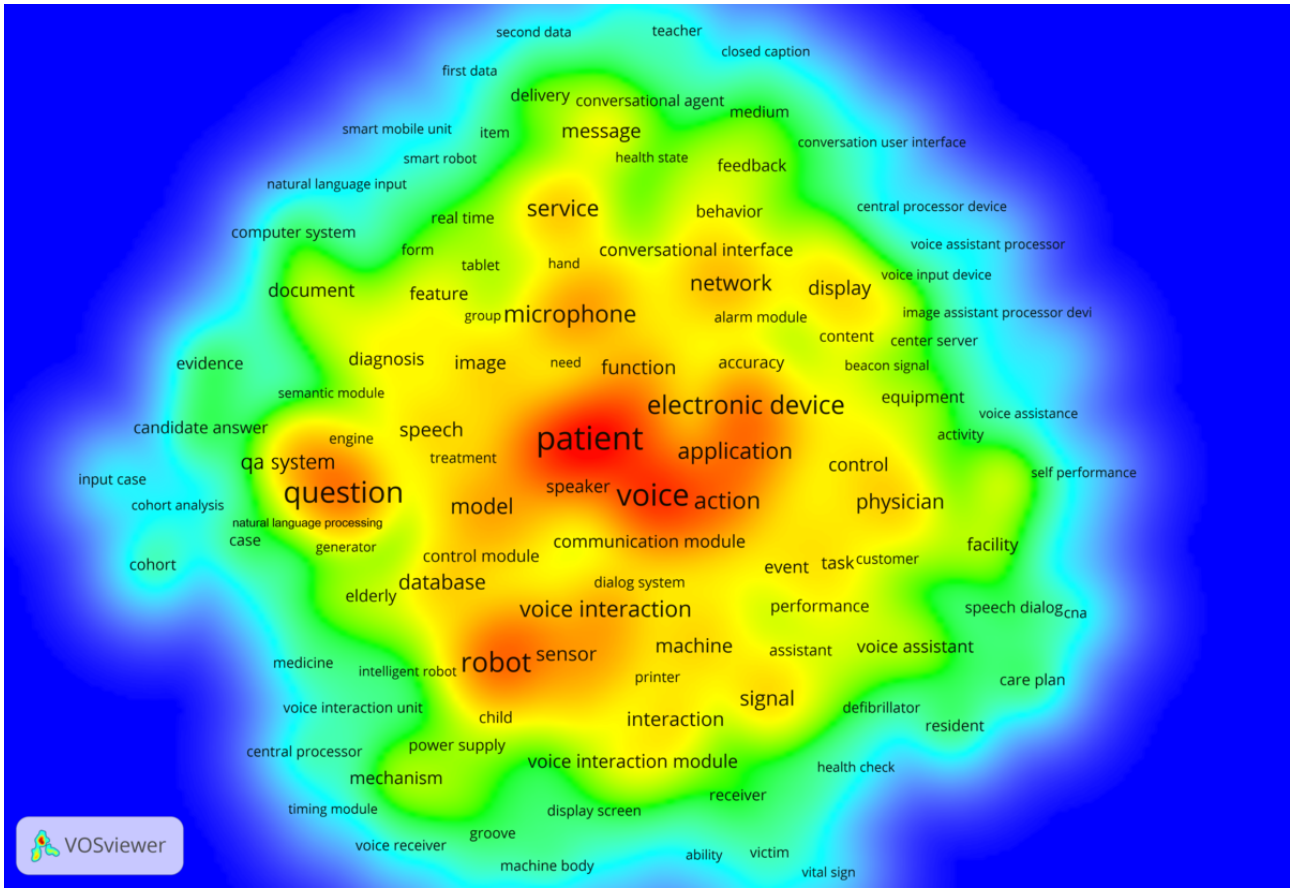
Figure 8. Heatmap of the topic terms that occurred in more than 9 publications.



Patents

The terms with top occurrence in patents (Figure 9) included *patient*, *voice*, and *electronic device*. Other key terms reflected the system components of CIs [56,57], such as *speaker* (18.5%, 30/162), *microphone* (15.4%, 25/162), and *communication module* (6.2%, 10/162). In addition, the terms *robot* (n_{patent}=34, 21.0%, 34/162), *sensor* (14.8%, 24/162), and *voice interaction* (42.6%, 69/162) were closely positioned and co-occurred frequently [58,59]. Another frequently occurred term was *question* (16%, 26/162), which also appeared closely with question answering systems, (9.3%, 15/162) [60,61]. The stakeholders addressed in the patents included not only *patient* but also *physician* (3.1%, 5/162) and *elderly* (2.5%, 4/162) [62,63].

Figure 9. Heatmap of the topic terms that occurred in more than 7 patents.



Discussion

Principal Findings

This is the first study to systematically examine emerging CI technologies for health purposes using grants, research publications, and patent data. We found increasing efforts in recent years in exploring health-related applications of CI.

There has been an increase in government funding support in this field in the United States and European countries since 2008. The total dollar amount of EC grants was higher than that of US grants. Research institutes and private organizations have been involved and awarded almost equally in European countries, whereas research institutions were the only type of awardee in the United States. These results may suggest that funding agencies in the United States need to encourage more collaboration between academia and industry.

The research publications are mainly from a few countries and institutions, and there is a lack of international collaborations across countries. US researchers have been leading in the number of publications about CIs for health. Northeastern University was the leading institute that has focused on the design and evaluation of CIs for health education and behavioral intervention in clinical settings, such as information access to clinical trials for cancer [39], depression therapy [64], or spinal cord injury recovery [65]. This university has established a wide research collaboration network, but its collaborators are mainly domestic institutions. Researchers in European countries produced about one-third of the included publications. CNRS, a prominent French research institute, has applied CIs to tackle health issues such as substance use disorder [66] and depression [67]. The collaboration in European countries is centered on France, and collaboration with US institutions is rarely observed, suggesting that CI research is still geographically isolated, and more cross-country collaboration could be encouraged. The Universiti Teknologi MARA system in Malaysia was the major contributor to research publications in Asia. Their research has focused on adopting the robot *NAO* in interventions and rehabilitation for patients with mental or brain disorders (eg, ASD [68] and cerebral palsy [69]). Osaka University in Japan had the most research collaborators compared with other Asian institutes. The institutions in Malaysia and China, on the other hand, were not actively involved in CI research collaboration. In addition to the lack of collaboration between countries, there is also a lack of collaboration between researchers and health care professionals. Several hospitals and medical centers engaged in interorganizational collaborations, but the number of such collaborations is still small. This may explain the existing challenge of insufficient clinical evaluation of CIs, inconsistent results, and high concern for patient safety, which were disclosed by previous studies [2,70,71]. In addition, the function of CIs in patient care requires substantial domain knowledge and an adequate understanding of human emotion [72]. Given these facts, the involvement of health care providers is important in design and clinical evaluation for better user-centered CI functions and improved patient outcomes.

The United States and China were the 2 most active countries in patenting activities, but the major players were different. In

the United States, private companies (eg, IBM, Vocollect Healthcare, and Google) have dominated patent applications, whereas in China, academic institutions are most actively engaged in patent applications. This finding is consistent with results from the WIPO's report on AI [25], which suggested that China has paid increasing attention to research-innovation translation. In addition, co-owned patents among university and companies were rarely identified in this study, which may suggest that academia-industry collaboration in patent activities related to CIs for health is currently uncommon, and further efforts are required to bridge this gap.

This study also presented key topic terms regarding the targeted users, health issues, and interaction modality. Overall, both *patient* and *robot* were addressed by all 3 types of records. *Child* and *elderly* appear to be 2 major user-related terms frequently discussed. For example, Northeastern University investigated computer-based ECA, which served as a humanoid assistant for patients with low literacy [73,74]. IBM's inventions, which primarily concerned clinical QA systems, helped physicians and other medical professionals to search medication evidence and supported their clinical decisions [60,61].

In terms of targeted health issues, this study found CIs were mainly used for addressing mental health and chronic diseases. The *asd* is the most frequently investigated term in grant applications and publications. It often occurred along with *children* and *robot*. This finding is consistent with the results from previous studies [2,75,76] that robots have been widely used to treat children with ASD by promoting their social behaviors and improving their communication skills. Our results also revealed that CIs were adopted to manage chronic diseases, such as stroke [46], cancer [47], and dementia [77]. These health issues usually require long-term treatment and intensive care. CIs, as a supplement to health care providers, can be used to promote communication and provide a support companion for patients with ASD [42,43] and to deliver interventions and management for patients with chronic diseases [48-51].

Regarding the human-CI interaction modality, we found that the CI interface was favorably presented as *robot* and that multimodality was widely discussed in publications and patents. For example, NAO robots interacted with users in both verbal and nonverbal modes and provided face-to-face communication and physical touch. More than 80 publications reported findings about the applications of this humanoid robot among patients with various health issues, including improving communication skills for children with ASD (eg, [42,43]), reducing distress for cancer patients (eg, [50]), and providing functional and emotional support for the elderly (eg, [52,78]). NAO's hybrid communication channels (eg, speech, eye contact, and body movement) contributed to users' in-depth engagement to achieve desired behavior change. This may suggest that the multimodal robot has become one of the main forms of CIs for health purposes. However, it does not mean that this is the only or best way to design or implement CIs for health. Instead, different approaches in presenting CIs for health can potentially benefit health care resolutions in different clinical scenarios [2,3]. Overall, empirical evidence has confirmed the positive effects of CIs' multimodal interface on medication adherence, social activity, and elderly patients' learning processes [3]. Nevertheless, there is a lack of consistency and evidence of long-term clinical effects of the CIs [75]. Therefore, future research should deepen the understanding of user interaction with multimodal CI agents and the corresponding impacts on patient outcomes.

Limitations

The study has several limitations. First, the selection of databases in this study introduced bias because we included only records released in English. Therefore, the results did not reflect any grants, research, and patenting activities reported in other languages. Second, during the data screening process, although the uncertain cases were reviewed and discussed by co-authors, only the first author screened the entire retrieved records, which may have led to bias in records inclusion and exclusion and consequently affected the results. Third, privately funded projects were not included in the analysis of grants data because of lack of access to private funding databases. Finally, the 3 types of data (ie, grants, publications, and patents) were retrieved and analyzed independently. There were no direct data mapping and integration across these data sources. Thus, results in this study should not be interpreted with respect to the relationship between public research investment, research outcomes, and patenting activities.

Conclusions

This study systematically analyzed CI technologies for health purposes using data extracted from multiple information sources. Our findings provided an overview of the countries, organizations, and topic terms in funding activities as well as the authorship, collaboration, contents, and related information of research publications and patents. Overall, the inclusion of grants and patents in addition to publications has presented complementary insights into the R&D landscape of the use of CIs for health purposes. Our results have shown that there is a lack of cross-sector collaboration among grantees in the United States. In addition, international collaboration in research should be encouraged among the institutes to sustain the growth of CI application in health. The academic-industrial collaboration should also be fostered in patenting activities. Although current CIs have focused extensively on mental health problems and chronic diseases, future research needs to extend CI technologies to more diverse health issues via safer and more engaging multimodal interfaces.

Acknowledgments

This work was supported by Health Science Library at the University of North Carolina at Chapel Hill, by R01NR016990 National Institute of Nursing Research (principal investigator: LS), by R21CA212516 National Cancer Institute (principal investigator: LS) and by University Cancer Research Fund of the Lineberger Comprehensive Cancer Center at the UNC-CH (principal investigator: LS).

Authors' Contributions

ZX developed the study design with inputs from FY, JD, and LS. ZX, FY, and JSW collected records and formulated the query with assistance from Jesse Akman and Lee Richardson in the Health Science Library at the University of North Carolina at Chapel Hill. ZX screened the title and abstract and resolved cases with uncertainty with FY and LS. ZX and FY prepared the initial draft of the manuscript. All authors provided contributions to the final version of the paper and approved it.

Conflicts of Interest

LS is an editorial board member for JMIR Aging.

Multimedia Appendix 1
Search query.
[DOCX File , 20 KB - jmir_v21i11e14672_app1.docx]

References

1. McTear M, Callejas Z, Griol D. Conversational interfaces: past and present. In: The Conversational Interface. Cham: Springer International Publishing; 2016:51-72.
2. Laranjo L, Dunn AG, Tong HL, Kocaballi AB, Chen J, Bashir R, et al. Conversational agents in healthcare: a systematic review. J Am Med Inform Assoc 2018 Sep 1;25(9):1248-1258 [FREE Full text] [doi: 10.1093/jamia/ocy072] [Medline: 30010941]
3. Montenegro JL, da Costa CA, Righi R. Survey of conversational agents in health. Expert Syst Appl 2019 Sep;129:56-67. [doi: 10.1016/j.eswa.2019.03.054]
4. Vaidyam AN, Wisniewski H, Halamka JD, Kashavan MS, Torous JB. Chatbots and conversational agents in mental health: a review of the psychiatric landscape. Can J Psychiatry 2019 Jul;64(7):456-464. [doi: 10.1177/0706743719828977] [Medline: 30897957]
5. Cohen PR, Oviatt SL. The role of voice input for human-machine communication. Proc Natl Acad Sci USA 1995 Oct 24;92(22):9921-9927 [FREE Full text] [doi: 10.1073/pnas.92.22.9921] [Medline: 7479803]
6. Cassell J. Embodied Conversational Agents. In: MIT press. Cambridge, MA: The MIT Press; 2000.
7. Oviatt S. Multimodal interfaces. In: Jacko JA, editor. Human Computer Interaction Handbook: Fundamentals, Evolving Technologies, and Emerging Applications. Second Edition. Boca Raton, Florida: CRC Press; 2007:286-304.
8. Bickmore TW, Mitchell SE, Jack BW, Paasche-Orlow MK, Pfeifer LM, Odonnell J. Response to a relational agent by hospital patients with depressive symptoms. Interact Comput 2010 Jul 1;22(4):289-298 [FREE Full text] [doi: 10.1016/j.intcom.2009.12.001] [Medline: 20628581]
9. Tanaka H, Negoro H, Iwasaka H, Nakamura S. Embodied conversational agents for multimodal automated social skills training in people with autism spectrum disorders. PLoS One 2017;12(8):e0182151 [FREE Full text] [doi: 10.1371/journal.pone.0182151] [Medline: 28796781]
10. Russo A, D'Onofrio G, Gangemi A, Giuliani F, Mongiovi M, Ricciardi F, et al. Dialogue systems and conversational agents for patients with dementia: the human-robot interaction. Rejuvenation Res 2019 Apr;22(2):109-120. [doi: 10.1089/rej.2018.2075] [Medline: 30033861]
11. Fitzpatrick KK, Darcy A, Vierhile M. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. JMIR Ment Health 2017 Jun 6;4(2):e19 [FREE Full text] [doi: 10.2196/mental.7785] [Medline: 28588005]
12. Brixey J, Hoegen R, Lan W, Rusow J, Singla K, Yin X. SHIHbot: A Facebook chatbot for Sexual Health Information on HIV/AIDS. In: Proceedings of the 18th Annual SIGdial Meeting on DiscourseDialogue Stroudsburg, PA. 2017 Aug Presented at: SIGDIAL'17; 2017; Saarbrücken, Germany p. 370-373. [doi: 10.18653/v1/W17-5544]
13. Cheng A, Raghavaraju V, Kanugo J, Handrianto Y, Shang Y. Development and Evaluation of a Healthy Coping Voice Interface Application Using the Google Home for Elderly Patients With Type 2 Diabetes. In: Proceedings of the 2018 15th IEEE Annual Consumer Communications & Networking Conference. 2018 Presented at: CCNC'18; January 12-15, 2018; Las Vegas, NV, USA. [doi: 10.1109/CCNC.2018.8319283]
14. Fadhil A. A conversational interface to improve medication adherence: towards AI support in patient's treatment. arXiv 2018:- [FREE Full text]
15. Fadhil A, Gabrielli S. Addressing Challenges in Promoting Healthy Lifestyles: the AI-Chatbot Approach. In: Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare. 2017 Presented at: PervasiveHealth'17; May 23-26, 2017; Barcelona, Spain p. 261-265. [doi: 10.1145/3154862.3154914]
16. Beveridge M, Fox J. Automatic generation of spoken dialogue from medical plans and ontologies. J Biomed Inform 2006 Oct;39(5):482-499 [FREE Full text] [doi: 10.1016/j.jbi.2005.12.008] [Medline: 16495159]
17. Barker DJ, van Schaik P, Simpson DS, Corbett WA. Evaluating a spoken dialogue system for recording clinical observations during an endoscopic examination. Med Inform Internet Med 2003 Jun;28(2):85-97. [Medline: 14692586]

18. Kenny PG, Parsons TD. Embodied conversational virtual patients. In: Perez-Marin D, Pascual-Nieto I, editors. Conversational Agents and Natural Language Interaction: Techniques and Effective Practices. Hershey, PA: IGI Global; 2011:254-281.
19. Lane J. Science innovation. Assessing the impact of science funding. Science 2009 Jun 5;324(5932):1273-1275. [doi: 10.1126/science.1175335] [Medline: 19498153]
20. Sanyal P. Understanding patents: The role Of R&D funding sources and the patent office. Econ Innov New Tech 2003;12(6):1. [doi: 10.1080/1043859031000078470]
21. Goeldner M, Herstatt C, Tietze F. The emergence of care robotics — A patent and publication analysis. Technol Forecast Soc Change 2015 Mar;92:115-131. [doi: 10.1016/j.techfore.2014.09.005]
22. Daim TU, Rueda G, Martin H, Gerdsri P. Forecasting emerging technologies: use of bibliometrics and patent analysis. Technol Forecast Soc Change 2006 Oct;73(8):981-1012. [doi: 10.1016/j.techfore.2006.04.004]
23. Chen H, Roco M, Son J, Jiang S, Larson C, Gao Q. Global nanotechnology development from 1991 to 2012: patents, scientific publications, and effect of NSF funding. J Nanopart Res 2013;15(9):1-21. [doi: 10.1007/s11051-013-1951-4]
24. Du J, Tang XL. Natural products against cancer: a comprehensive bibliometric study of the research projects, publications, patents and drugs. J Cancer Res Ther 2014 Aug;10(Suppl 1):27-37 [FREE Full text] [doi: 10.4103/0973-1482.139750] [Medline: 25207887]
25. WIPO. WIPO Technology Trends 2019 Artificial Intelligence. Geneva, Switzerland: World Intellectual Property Organization; 2019. URL: https://www.wipo.int/publications/en/details.jsp?id=4386
26. Gao J, Galley M, Li L. Neural Approaches to Conversational AI. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics: Tutorial Abstracts. 2018 Presented at: ACL'18; July 16th–18th, 2018 ; Melbourne, Australia p. 1371-1374. [doi: 10.1145/3209978.3210183]
27. Wohlin C. Guidelines for Snowballing in Systematic Literature Studies and a Replication in Software Engineering. In: Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering. 2014 Presented at: EASE'14; May 13-14, 2014; London, England, United Kingdom. [doi: 10.1145/2601248.2601268]
28. Lv PH, Wang G, Wan Y, Liu J, Liu Q, Ma F. Bibliometric trend analysis on global graphene research. Scientometrics 2011;88(2):399-419. [doi: 10.1007/s11192-011-0386-x]
29. Narin F. Patent bibliometrics. Scientometrics 1994 May;30(1):147-155. [doi: 10.1007/bf02017219]
30. Guan J, He Y. Patent-bibliometric analysis on the Chinese science — technology linkages. Scientometrics 2007;72(3):403-425. [doi: 10.1007/s11192-007-1741-1]
31. Otte E, Rousseau R. Social network analysis: a powerful strategy, also for the information sciences. J Inf Sci 2002 Oct;28(6):441-453. [doi: 10.1177/016555150202800601]
32. Perianes-Rodríguez A, Olmeda-Gómez C, Moya-Anegón F. Detecting, identifying and visualizing research groups in co-authorship networks. Scientometrics 2010;82(2):307-319. [doi: 10.1007/s11192-009-0040-z]
33. Callon M, Courtial J, Turner WA, Bauin S. From translations to problematic networks: an introduction to co-word analysis. Soc Sci Inf 1983 Mar 1;22(2):191-235. [doi: 10.1177/053901883022002003]
34. van Eck NJ, Waltman L. Visualizing bibliometric networks. In: Ding Y, Rousseau R, Wolfram D, editors. Measuring Scholarly Impact: Methods and Practice. Cham: Springer; 2014:285-320.
35. Radhakrishnan S, Erbis S, Isaacs JA, Kamarthi S. Novel keyword co-occurrence network-based methods to foster systematic reviews of scientific literature. PLoS One 2017;12(3):e0172778 [FREE Full text] [doi: 10.1371/journal.pone.0172778] [Medline: 28328983]
36. United States Patent and Trademark Office. 211 Claiming the Benefit of an Earlier Filing Date Under 35 USC 120 and 119(e) [R-07.2015]. URL: https://www.uspto.gov/web/offices/pac/mpep/s211.html [accessed 2019-04-19] [WebCite Cache ID 77kennk5S]
37. van Eck NJ, Waltman L. Software survey: VOSviewer, a computer program for bibliometric mapping. Scientometrics 2010 Aug;84(2):523-538 [FREE Full text] [doi: 10.1007/s11192-009-0146-3] [Medline: 20585380]
38. van Eck NJ, Waltman L. Text mining and visualization using VOSviewer. arXiv preprint arXiv:1109.2058 2011 Sep:1-5 [FREE Full text]
39. Bickmore TW, Utami D, Matsuyama R, Paasche-Orlow MK. Improving access to online health information with conversational agents: a randomized controlled experiment. J Med Internet Res 2016 Jan 4;18(1):e1 [FREE Full text] [doi: 10.2196/jmir.5239] [Medline: 26728964]
40. Greenwald JL, Jack BW. Preventing the preventable: reducing rehospitalizations through coordinated, patient-centered discharge processes. Prof Case Manag 2009;14(3):135-40; quiz 141 [FREE Full text] [doi: 10.1097/NCM.0b013e318198d4e1] [Medline: 19474639]
41. Burton C, Tatar AS, McKinstry B, Matheson C, Matu S, Moldovan R, Help4Mood Consortium. Pilot randomised controlled trial of Help4Mood, an embodied virtual agent-based system to support treatment of depression. J Telemed Telecare 2016 Sep;22(6):348-355. [doi: 10.1177/1357633X15609793] [Medline: 26453910]
42. Ioannou A, Kartapanis I, Zaphirisocial P. Social robots as co-therapists in autism therapy sessions: a single-case study. In: Tapus A, André E, Martin JC, Ferland F, Ammi M, editors. Social Robotics. Cham: Springer; 2015:255-263.

43. Baraka K, Melo F, Veloso M. 'Autistic robots' for embodied emulation of behaviors typically seen in children with different autism severities. In: Kheddar A, Yoshida E, Ge SS, Suzuki K, Cabibihan JJ, Eyssel F, editors. Social Robotics. Cham: Springer International Publishing; 2017:105-114.
44. Kuwabara K, Shimode Y, Miyamoto S. Agent-based Remote Conversation Support for People With Aphasia. In: Proceedings of the 4th KES international conference on Agent and multi-agent systems: technologies and applications. 2010 Presented at: KES-AMSTA'10; June 23-25, 2010; Gdynia, Poland p. 371-380. [doi: 10.1007/978-3-642-13480-7_39]
45. Kuwabara K, Hayashi S, Uesato T, Umadome K, Takenaka K. Remote conversation support for people with aphasia: some experiments and lessons learned. In: Stephanidis C, editor. Universal Access in Human-Computer Interaction. Addressing Diversity. Berlin, Heidelberg: Springer; 2009:375-384.
46. Malik NA, Yussof H, Hanapiah FA, Rahman RA, Basri HH. Human-robot interaction for children with cerebral palsy: reflection and suggestion for interactive scenario design. Procedia Comput Sci 2015;76:388-393. [doi: 10.1016/j.procs.2015.12.315]
47. Borovac B, Gnjatovic M, Savic S, Rakovic M, Nikolic M. Human-like robot MARKO in the rehabilitation of children with cerebral palsy. In: Wenger P, Chevallereau C, Pisla D, Bleuler H, Rodic A, editors. New Trends in Medical and Service Robots: Human Centered Analysis, Control and Design. Cham: Springer; 2016:191-203.
48. Zannatha JM, Tamayo AJ, Sánchez AD, Delgado JE, Cheu LE, Arévalo WA. Development of a system based on 3D vision, interactive virtual environments, ergonometric signals and a humanoid for stroke rehabilitation. Comput Methods Programs Biomed 2013 Nov;112(2):239-249. [doi: 10.1016/j.cmpb.2013.04.021] [Medline: 23827333]
49. Jung H, Takahashi T, Choe Y, Baird J, Foster T, Grupen R. Towards extended virtual presence of the therapist in stroke rehabilitation. IEEE Int Conf Rehabil Robot 2013 Jun;2013:6650345. [doi: 10.1109/ICORR.2013.6650345] [Medline: 24187164]
50. Alemi M, Ghanbarzadeh A, Meghdari A, Moghadam LJ. Clinical application of a humanoid robot in pediatric cancer interventions. Int J of Soc Robotics 2016;8(5):743-759. [doi: 10.1007/s12369-015-0294-y]
51. Hassoon A, Schrack J, Naiman D, Lansey D, Baig Y, Stearns V, et al. Increasing physical activity amongst overweight and obese cancer survivors using an Alexa-based intelligent agent for patient coaching: protocol for the physical activity by technology help (PATH) trial. JMIR Res Protoc 2018 Feb 12;7(2):e27 [FREE Full text] [doi: 10.2196/resprot.9096] [Medline: 29434016]
52. Liu T, Meng M. Study on cooperation between humanoid robot Nao and Barrett WAM. In: 2012 IEEE International Conference on Robotics and Biomimetics. 2012 Presented at: ROBIO'12; December 11-14, 2012; Guangzhou, China p. 980-983. [doi: 10.1109/robio.2012.6491096]
53. Carelli L, Gaggioli A, Pioggia G, De Rossi F, Riva G. Affective robot for elderly assistance. Stud Health Technol Inform 2009;144:44-49. [Medline: 19592728]
54. Burger B, Ferrané I, Lerasle F, Infantes G. Two-handed gesture recognition and fusion with speech to command a robot. Auton Robot 2012;32(2):129-147. [doi: 10.1007/s10514-011-9263-y]
55. Garzotto F, Gelsomini M, Kinoe Y. Puffy: A mobile inflatable interactive companion for children with neurodevelopmental disorder. In: Bernhaupt R, Dalvi G, Joshi A, Balkrishan D, O'Neill J, Winckler M, editors. Human-Computer Interaction - INTERACT 2017. Cham: Springer; 2017:467-492.
56. Wang F, Chen L, Qu W, Zhu Y, Inventors. Elderly care service robot. Chinese Patent CN201810749966.3A 2018 Jul 10:NA-NA.
57. Fountaine D. Google Patents. Voice Controlled Assistance for Monitoring Adverse Events of a User and/or Coordinating Emergency Actions Such as Caregiver Communication. URL: https://patents.google.com/patent/US10258295B2/en [accessed 2019-11-05]
58. Wu K, Wu L, Cui H, Inventor. Health-service robot capable of achieving electrocardiograph detection. Chinese Patent CN105310686A 2015 Aug 24:NA-NA.
59. Chen M, Tian Q, Li X, Zhang L, Inventor. Intelligent limb motion monitor and application method. Chinese Patent CN106267730A 2016 Aug 12:NA-NA.
60. Kozloski J, Murdock J, Pickover P, Walker G. Google Patents. Empathy Injection for Question-Answering Systems. URL: https://patents.google.com/patent/US9298766B2/en [accessed 2019-11-05]
61. Anumalasetty K, Ayyagari P, Bhide M. Patent Scope. Personal Assistant Computing System Monitoring. URL: https://patentscope.wipo.int/search/en/detail.jsf;jsessionid=3EAB4FBD23824DC2313883CD6F122CDE.wapp2nC?docId=US240620712&recNum=1080&office=&queryString=&prevFilter=&sortOption=Pub+Date+Desc&maxRec=74636364 [accessed 2019-11-05]
62. Bagchi S, Ferrucci D, Levas A, Mueller E. Patent Scope. A Decision-Support Application and System for Problem Solving Using a Question-Answering System. URL: https://patentscope.wipo.int/search/en/detail.jsf?docId=WO2012122198&recNum=184&docAn=US2012027942&queryString=((num& [accessed 2019-11-05]
63. Kawahara T, Watabe S, Nakano Y. Voice interactive method and voice interactive system. Japan Patent JP2016038501A 2014 Aug 08:NA-NA.
64. Alemi M, Meghdari A, Ghanbarzadeh A, Moghadam L, Ghanbarzadeh A. Effect of Utilizing a Humanoid Robot as a Therapy-assistant in Reducing Anger, Anxiety, and Depression. In: Proceedings of the 2014 Second RSI/ISM International

Conference on Robotics and Mechatronics. 2014 Presented at: ICRoM'14; October 15-17, 2014; Tehran, Iran p. 748-753. [doi: 10.1109/icrom.2014.6990993]
65. Shamekhi A, Trinh H, Bickmore T, DeAngelis T, Ellis T, Houlihan B. A Virtual Self-care Coach for Individuals with Spinal Cord Injury. In: Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility. Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility - ASSETS" ?16 New York, New York, USA; 2016 Presented at: ASSETS'16; October 24-26, 2016; Reno, Nevada, USA p. 327-328. [doi: 10.1145/2982142.2982199]
66. Auriacombe M, Moriceau S, Serre F, Denis C, Micoulaud-Franchi J, de Sevin E, et al. Development and validation of a virtual agent to screen tobacco and alcohol use disorders. Drug Alcohol Depend 2018 Dec 1;193:1-6. [doi: 10.1016/j.drugalcdep.2018.08.025] [Medline: 30321739]
67. Philip P, Micoulaud-Franchi J, Sagaspe P, Sevin ED, Olive J, Bioulac S, et al. Virtual human as a new diagnostic tool, a proof of concept study in the field of major depressive disorders. Sci Rep 2017 Feb 16;7:42656 [FREE Full text] [doi: 10.1038/srep42656] [Medline: 28205601]
68. Aziz AA, Moganan FF, Ismail A, Lokman AM. Autistic children's kansei responses towards humanoid-robot as teaching mediator. Procedia Comput Sci 2015;76:488-493. [doi: 10.1016/j.procs.2015.12.322]
69. Abdul MN, Yussof H, Hanapiah F, Anne S. Human Robot Interaction (HRI) between a humanoid robot and children with Cerebral Palsy: Experimental framework and measure of engagement. In: Proceedings of the 2014 IEEE Conference on Biomedical Engineering and Sciences. 2014 Presented at: IECBES'14; December 8-10, 2014; Kuala Lumpur, Malaysia p. 430-435. [doi: 10.1109/iecbes.2014.7047536]
70. Bickmore T, Trinh H, Asadi R, Olafsson S. Safety first: conversational agents for health care. In: Moore RJ, Szymanski MH, Arar R, Ren GJ, editors. Studies in Conversational UX Design. Cham: Springer; 2018:33-57.
71. Miner AS, Milstein A, Schueller S, Hegde R, Mangurian C, Linos E. Smartphone-based conversational agents and responses to questions about mental health, interpersonal violence, and physical health. JAMA Intern Med 2016 May 1;176(5):619-625 [FREE Full text] [doi: 10.1001/jamainternmed.2016.0400] [Medline: 26974260]
72. Palanica A, Flaschner P, Thommandram A, Li M, Fossat Y. Physicians' perceptions of chatbots in health care: cross-sectional web-based survey. J Med Internet Res 2019 Apr 5;21(4):e12887 [FREE Full text] [doi: 10.2196/12887] [Medline: 30950796]
73. Bickmore TW, Pfeifer LM, Paasche-Orlow MK. Using computer agents to explain medical documents to patients with low health literacy. Patient Educ Couns 2009 Jun;75(3):315-320 [FREE Full text] [doi: 10.1016/j.pec.2009.02.007] [Medline: 19297116]
74. Bickmore T, Pfeifer L, Jack B. Taking the Time to Care: Empowering Low Health Literacy Hospital Patients With Virtual Nurse Agents. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2009 Presented at: CHI'09; April 4–9, 2009; Boston, MA, USA p. 1265-1274. [doi: 10.1145/1518701.1518891]
75. Provoost S, Lau HM, Ruwaard J, Riper H. Embodied conversational agents in clinical psychology: a scoping review. J Med Internet Res 2017 May 9;19(5):e151 [FREE Full text] [doi: 10.2196/jmir.6553] [Medline: 28487267]
76. Pennisi P, Tonacci A, Tartarisco G, Billeci L, Ruta L, Gangemi S, et al. Autism and social robotics: a systematic review. Autism Res 2016 Feb;9(2):165-183. [doi: 10.1002/aur.1527] [Medline: 26483270]
77. Sarabia M, Young N, Canavan K, Edginton T, Demiris Y, Vizcaychipi MP. Assistive robotic technology to combat social isolation in acute hospital settings. Int J of Soc Robotics 2018;10(5):607-620. [doi: 10.1007/s12369-017-0421-z]
78. Yu N, Xu C, Wang K, Yang Z, Liu J. Gesture-based Telemanipulation of a Humanoid Robot for Home Service Tasks. In: Proceedings of the 2015 IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems. 2015 Presented at: CYBER'15; June 8-12, 2015; Shenyang, China p. 1923-1927. [doi: 10.1109/cyber.2015.7288241]

Abbreviations

AHRQ: Agency for Healthcare Research and Quality
ASD: autism spectrum disorder
CIs: conversational interfaces
CNRS: National Center for Scientific Research
CORDIS: Community Research and Development Information Service
DII: Derwent Innovation Index
EC: Europe Commission
EU: European Union
NIH: National Institutes of Health
NSF: National Science Foundation
PI: principal investigator
R&D: research and development
WIPO: World Intelligence Property Organization

Edited by G Eysenbach; submitted 13.05.19; peer-reviewed by S Zhou, D Carvalho; comments to author 28.05.19; revised version received 10.10.19; accepted 19.10.19; published 18.11.19.

Please cite as:
Xing Z, Yu F, Du J, Walker JS, Paulson CB, Mani NS, Song L
Conversational Interfaces for Health: Bibliometric Analysis of Grants, Publications, and Patents
J Med Internet Res 2019;21(11):e14672
URL: http://www.jmir.org/2019/11/e14672/
doi:10.2196/14672
PMID:31738171

Original Paper

A Chatbot Versus Physicians to Provide Information for Patients With Breast Cancer: Blind, Randomized Controlled Noninferiority Trial

Jean-Emmanuel Bibault[1*], MD, PhD; Benjamin Chaix[2,3*], MSc; Arthur Guillemassé[3], MSc; Sophie Cousin[4], MSc, MD; Alexandre Escande[5], MSc, MD; Morgane Perrin[6], MSc, MD; Arthur Pienkowski[3], PharmD; Guillaume Delamon[3], PharmD; Pierre Nectoux[3], MSc; Benoît Brouard[3], PharmD

[1]Department of Radiation Oncology, Hôpital Européen Georges Pompidou, AP-HP, Paris, France
[2]ENT Department, Hôpital Gui de Chauliac, Université Montpellier 1, Montpellier, France
[3]Wefight, Institut du Cerveau et de la Moelle épinière, Hôpital Pitié-Salpêtrière, Paris, France
[4]Department of Medical Oncology, Institut Bergonié, Bordeaux, France
[5]Department of Radiation Oncology, Centre Oscar Lambret, Lille, France
[6]Department of Gynecological Oncologic Surgery, Gustave Roussy Cancer Campus, Villejuif, France
[*]these authors contributed equally

Corresponding Author:
Benjamin Chaix, MSc
ENT Department
Hôpital Gui de Chauliac
Université Montpellier 1
60 Avenue Augustin Fliche
Montpellier, 34264
France
Phone: 33 0467336872
Email: b-chaix@chu-montpellier.fr

Abstract

Background: The data regarding the use of conversational agents in oncology are scarce.

Objective: The aim of this study was to verify whether an artificial conversational agent was able to provide answers to patients with breast cancer with a level of satisfaction similar to the answers given by a group of physicians.

Methods: This study is a blind, noninferiority randomized controlled trial that compared the information given by the chatbot, Vik, with that given by a multidisciplinary group of physicians to patients with breast cancer. Patients were women with breast cancer in treatment or in remission. The European Organisation for Research and Treatment of Cancer Quality of Life Group information questionnaire (EORTC QLQ-INFO25) was adapted and used to compare the quality of the information provided to patients by the physician or the chatbot. The primary outcome was to show that the answers given by the Vik chatbot to common questions asked by patients with breast cancer about their therapy management are at least as satisfying as answers given by a multidisciplinary medical committee by comparing the success rate in each group (defined by a score above 3). The secondary objective was to compare the average scores obtained by the chatbot and physicians for each INFO25 item.

Results: A total of 142 patients were included and randomized into two groups of 71. They were all female with a mean age of 42 years (SD 19). The success rates (as defined by a score >3) was 69% (49/71) in the chatbot group versus 64% (46/71) in the physicians group. The binomial test showed the noninferiority (P<.001) of the chatbot's answers.

Conclusions: This is the first study that assessed an artificial conversational agent used to inform patients with cancer. The EORTC INFO25 scores from the chatbot were found to be noninferior to the scores of the physicians. Artificial conversational agents may save patients with minor health concerns from a visit to the doctor. This could allow clinicians to spend more time to treat patients who need a consultation the most.

Trial Registration: Clinicaltrials.gov NCT03556813, https://tinyurl.com/rgtlehq

(J Med Internet Res 2019;21(11):e15787) doi:10.2196/15787

KEYWORDS

chatbot; clinical trial; cancer

Introduction

Background

Chatbots can imitate human conversation by using a field of artificial intelligence (AI) known as natural language processing. Chatbots are now widely used in several forms as voice-based agents, such as Siri (Apple), Google Now (Google), Alexa (Amazon), or Cortana (Microsoft). Text-based chatbots are available as Messenger (Facebook) agents or as stand-alone mobile or Web apps. They provide information and create a dynamic interaction between the agent and the user, without human back-end intervention. The concept of an artificial conversational agent dates back to 1950, when Alan Turing envisioned a future where a computer would be able to express itself with a level of sophistication that would render it indistinguishable from humans [1].

In health care, the first example of a computer program used as a conversational agent was Joseph Weizenbaum's ELIZA, a program that mimicked a Rogerian psychotherapist and that was able to rephrase the patient's sentences as questions and provide prerecorded answers [2]. In 1991, Dr Sbaitso was created as an AI speech synthesis program for MS-DOS personal computers. In this software, Dr Sbaitso was designed as a psychologist, with very limited possibilities [3]. Four years later, the chatbot, Artificial Linguistic Internet Computer Entity, was created to include 40,000 knowledge categories and was awarded the Loebner Prize thrice [4]. In 2001, SmarterChild was made available as a bot distributed across SMS networks and is now considered as a precursor to Apple's Siri, which was released on iPhones in 2010. Patients can now use chatbots to check for symptoms and to monitor their health, but the relevance and validity of chatbots have rarely been assessed [5-7].

Objective

Wefight designed a chatbot named Vik for patients with breast cancer and their relatives via personalized text messages. Vik provides information about breast cancer and its epidemiology, treatments, side effects, and quality of life improvement strategies (sport, fertility, sexuality, and diet). More practical information, such as reimbursement and patients' rights, is also available. Chaix et al [8] showed that it was possible to obtain support through a chatbot as Vik improved the medication adherence rate of patients with breast cancer. Vik is available for free on the Web, on any mobile phones, iOS (Apple) or Android (Google), or on Messenger (Facebook).

This study is a blind, noninferiority randomized controlled trial that compared the information given by the Vik chatbot versus that given by a multidisciplinary group of physicians (medical, radiation, and surgical oncology) to patients with breast cancer (NCT03556813). The EORTC QLQ-INFO25 questionnaire, which was validated to assess information of patients with cancer [9], was adapted and used to compare the quality of the information provided to the 2 groups of patients by the physician or the chatbot.

Methods

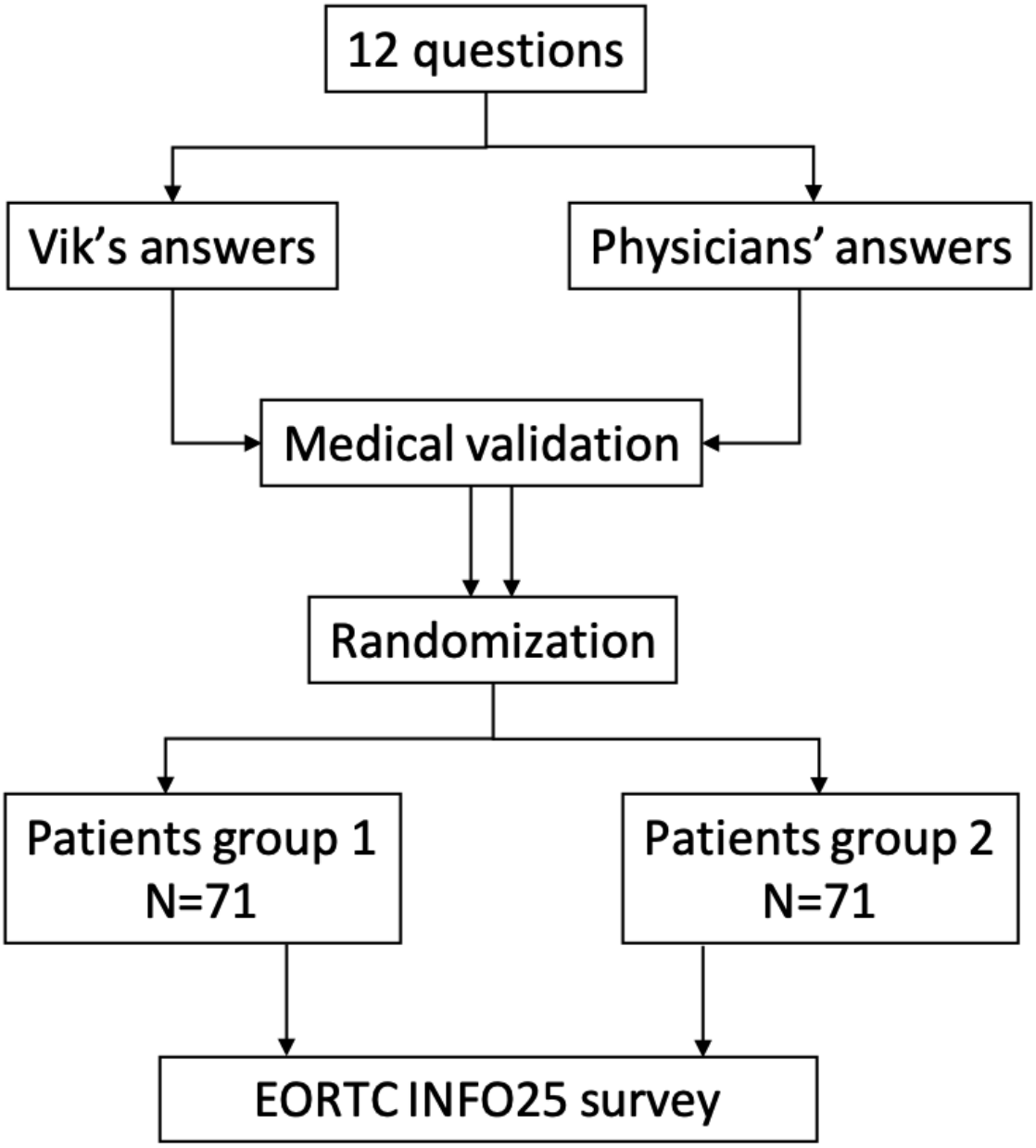
Study Design and Participants

The study was a blind, noninterventional, noninferiority randomized study, without any risk or burden. It was conducted in France in November and December 2018.

The authors selected the 12 most frequently asked questions about breast cancer from Vik's database (Multimedia Appendix 1). These questions were then asked both to the Vik chatbot and to a multidisciplinary medical committee (oncologist surgeon, medical oncologist, and oncologist radiotherapist; Figure 1). The second independent multidisciplinary group of physicians ensured that each group's answers did not provide inaccurate information. Institutional affiliations of the coordinating team were not displayed.

Patients were recruited with the help of a French breast cancer patients association (Mon Réseau Cancer du Sein). They were filtered for eligibility based on the inclusion criteria (age >18 years, female, subjects with breast cancer in treatment or remission, nonopposition, and internet literacy). Participants were compensated for their time. This study was approved by an ethics committee independently selected by the French Ministry of Health (N° ID RCB: 2018-A01365-50) and registered in the ClinicalTrials.gov database (NCT03556813). The data collected were anonymized and then hosted by Wefight on a server compliant with health care data storage requirements. Consent was collected online before the start of the study. In accordance with the French and European laws on information technology and civil liberties (Commission Nationale Informatique et Libertés, Règlement Général pour la Protection des Données), users had a right of use at their disposal to verify its accuracy and, if necessary, to correct, complete, and update it. They also had a right to object to their use and a right to delete these data. General conditions of use were displayed and explained very clearly, and they must be accepted before using the questionnaire. No demographical data beyond age were asked or gathered to participate in the study.

Figure 1. Flow diagram. EORTC QLQ-INFO25: European Organisation for Research and Treatment of Cancer.



Chatbot Design

Wefight designed a chatbot named Vik to empower patients with cancer and their relatives via personalized text messages. Vik's answers are very diverse, and patients can find all the relevant, quality-checked medical information they need. Vik's architecture is composed of several technological parts allowing a fine analysis of the questions posed by the patients and an adapted treatment of the answer.

For a chatbot to be fully developed, both machine learning algorithms and natural language processing are required. To build a chatbot, there are 2 crucial components to be supervised: intent classification and entity recognition. To understand the users' messages and send personalized answers, the conversation goes through 3 steps: the first step analyzes the sentence and identifies intents and entities by using machine learning. The second stage activates modules according to the intents and entities detected by the first stage, and the third stage aggregates the answers of all activated modules to build the answer sent to the user and saves the conversation on the user's profile.

For the patient, the use of a chatbot is very simple. It is a classic chat on a conversation window. The patient asks a question by writing it on his or her keyboard, and the chatbot answers directly in simple and understandable language.

Procedures

Patients were randomized (1:1) blindly and received either the responses of the Vik chatbot or the responses of the medical committee to the predefined 12 questions, as previously explained. Participants were shown each question in order, and blinded responses were directly delivered as Web-based text messages for each group. The full answer for each question from either Vik or experts was directly shown to the participants upon activation of each question by the participants. There was no actual conversation per question nor the necessity of natural language processing for each question. Patients were then asked

to complete an adapted version of the EORTC QLQ-INFO25 questionnaire online, assessing the quality of medical information received based on each response. A total of 21 items of the EORTC QLQ-INFO25 questionnaire were included (Multimedia Appendix 1).

Outcomes

The perceived quality of the answers was assessed using the QLQ-INFO25 questionnaire that uses a scale of satisfaction graded from 1 to 4.

The primary objective was to assess the overall perceived quality of the answers given by the Vik chatbot to common questions asked by patients with breast cancer about their therapy management compared with answers given by a multidisciplinary medical committee (oncologist surgeon, medical oncologist, and radiotherapist oncologist), by comparing proportions of success in the physicians' and Vik's group. The secondary objective was to compare the average scores obtained by the chatbot and by the physicians for each individual INFO25 item. Gradings for the 21 items were averaged to define an overall score for each patient, in each group. We defined success as a grade greater than or equal to 3. Descriptive statistics were used to summarize patient characteristics by treatment group.

Statistical Analysis

This study used a randomized phase III design with an alpha of .05 and a beta of .2 with a noninferiority limit of 10%. The effect size was based on a published EORTC INFO25 validation study [9]. This noninferiority limit of 10% was chosen as an acceptable difference for patient satisfaction. In view of these assumptions, the trial required at least 142 patients randomly assigned to the 2 groups. A 1-sided binomial test using the method of Mietinen and Nurminen was performed to compare the difference between the proportions of success in the 2 groups for questions 1 to 19 and the noninferiority limit. Noninferiority was declared if the *P* value of the test is lower than .05. For each item, confidence interval of the difference between the proportions of success in the physicians' group and Vik's group was estimated using the Wald Z method. Noninferiority was declared when the upper limit of a 2-sided 90% CI, equivalent to a 1-sided 95% CI, did not exceed the noninferiority limit of 10%.

Results

Analysis Size

Between November and December 2018, we included a total of 142 patients, divided into 2 groups of 71. For each group, the numbers of participants who were randomly assigned received the intended treatment and were analyzed for the primary outcome. A single intervention was performed for this study. All participating patients finished the evaluation. They were all female with a mean age of 42 years (SD 19).

Descriptive Analysis

Patients responded to the questionnaire in an average of 15 min (SD 4). The first group of 71 patients received the responses from Vik, and the second group received the responses from physicians. The average global rating was 2.86 (median 3, IQR 2-4). The success rates (as defined by a score >3) were 69% in the chatbot group versus 64% in the physicians group. Patients assessing physicians' answers gave an average rating of 2.82, whereas patients assessing Vik's answers gave an average rating of 2.89 (Multimedia Appendix 2).

A total of 62.0% of patients (88/142) would have liked to get even more information (65% [46/71]) in the physicians' group and 59% ([42/71] in Vik's group), whereas only 4.2% (6/142) would have liked to get less. A total of 83.1% of patients (118/142) found answers helpful (82% [58/71] in the physicians' group and 85% [60/71] in Vik's group), and 81.0% (115/142) were satisfied with the amount of information they have received (77% [55/71] in physicians' group and 85% [60/71] in Vik's group).

Comparison of Patient Groups

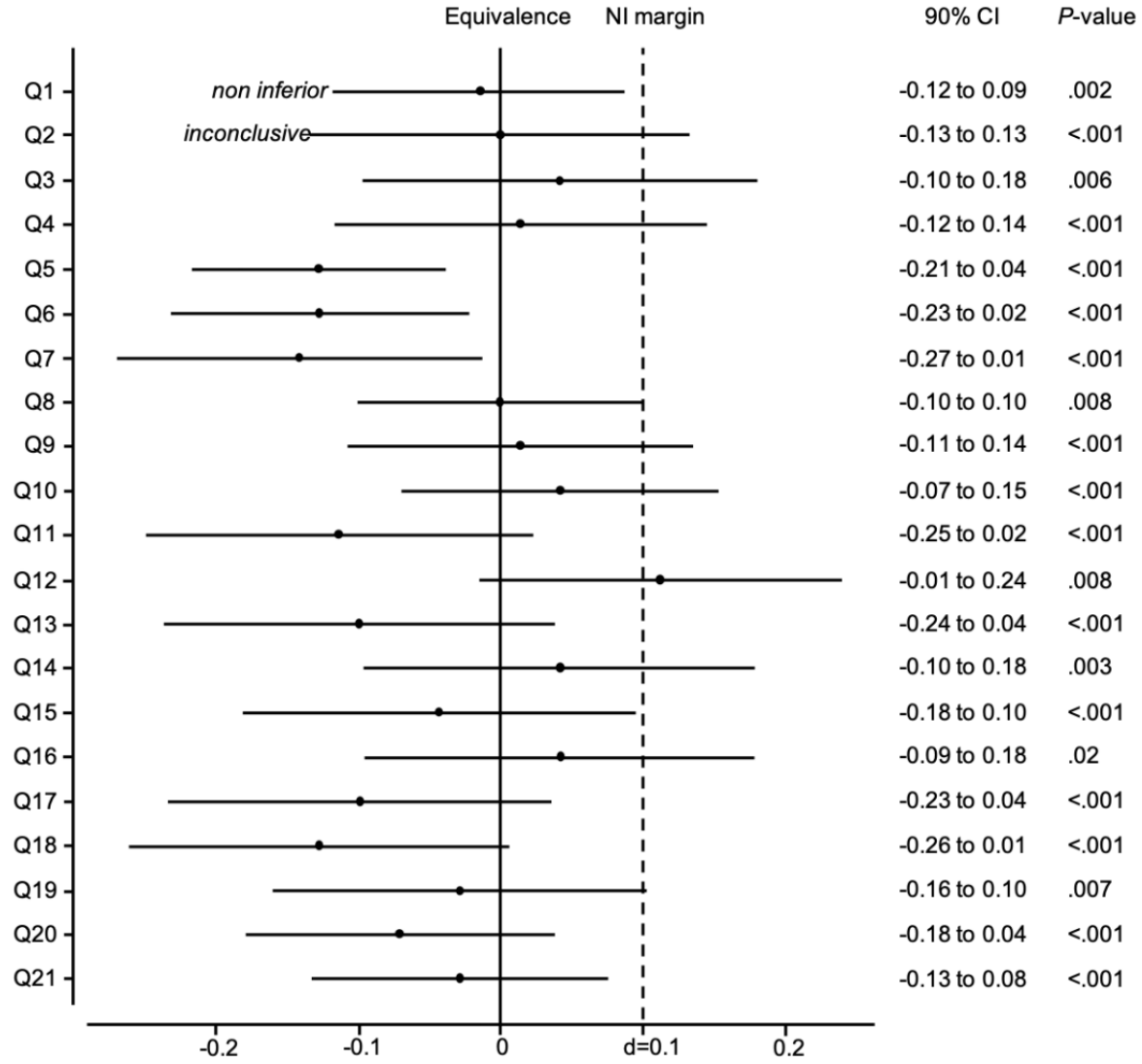
Primary Objective

The difference between success rates in the physicians' group and Vik's group was –0.03 (95% CI -0.07 to 0.00). Furthermore, the binomial test showed a noninferiority (P<1e-14) between the perceived quality of the chatbot responses and that of the physicians, as assessed by EORTC INFO25.

Secondary Objective

Both-sided 90% CI, equivalent to 1-sided 95% CI was computed for the difference between proportions of success in the physicians' group and Vik's group for each item (Multimedia Appendix 2). For 12 items of them, the noninferiority can be declared as the upper limit of the 95% CI did not exceed the 0.1 noninferiority limit (Figure 2). For the rest of them (9 items), the upper limit of the 95% CI crossed the 0.1 noninferiority limit. For these items, the noninferiority cannot be claimed. These items include questions 2 and 3 about breast cancer stages and causes, question 4 about whether or not the cancer is under control, 4 questions related to treatments (types, benefits, and side effects), and 2 questions related to care outside of the hospital.

Figure 2. Noninferiority (NI) graph.



Discussion

Principal Findings

This is the first study that rigorously assessed an artificial conversational agent used to inform patients with cancer, but the study has limitations: we did not evaluate the demographic features of the patients who answered the survey to remain in compliance with the European General Data Protection Regulation. Patients were recruited in our study through a patients association mailing list, which means that they could potentially be younger than the average population of patients with breast cancer, have more digital literacy skills, and be more open-minded toward digital tools, even if the 2 groups were blinded and did not know if they received the answers from the chatbot or from the group of physicians.

Chatbot Assessment

A search on ClinicalTrials.gov currently returns only 4 trials evaluating chatbots in health care: in the United Kingdom, a nonrandomized trial is being performed by the National Health Service to compare the Babylon chatbot with the nonemergency 111 telephone number [10]. Patients interact with an automatic agent to describe their symptoms. Advices and information are given in return by the chatbot. The second trial, The Buddy Study (NCT02742740) evaluates an Embodied Conversational Agent (ECA) Oncology Trial Advisor for Cancer Trials that acts as an advisor to patients on chemotherapy regimens, promoting protocol adherence and retention, providing anticipatory guidance, and answering questions. The chatbot also serves as a conduit to capture information about complaints or adverse events. Usability metrics will include session time, satisfaction, and error rates. Subjects will be identified from among patients on chemotherapy regimens at the Boston Medical Center [11]. All subjects will be enrolled for 2 months and randomized to the chatbot group or control group. The primary outcome will be treatment protocol adherence, defined by the number of treatment visits attended/number of treatment visits scheduled. The secondary outcome will measure subject satisfaction, number of adverse events as reported through the ECA and directly to clinic by patient, time to detect and resolve adverse events as reported through the ECA and directly to clinic by patient, and adverse event false alarm rate as reported through ECA and directly to clinic by patient. The third study, the RAISE project (NCT01458002) [12], is designed to promote exercise and sun protection. The primary aims were to develop

and assess the effectiveness of a tailored internet intervention on a national sample, to develop and assess the effectiveness of the internet intervention enhanced by a relational agent, and to determine if the intervention with the relational agent can outperform the regular tailored internet intervention. The study will include 3 groups (control, internet, and internet plus relational agent). A representative national sample of 1639 individuals at risk for both behaviors will be recruited.

Randomized studies demonstrating the superiority (or at least noninferiority) of chatbots, compared with an intervention performed by a physician, do not exist. However, if chatbots are to be safely used by a large number of patients, they must be evaluated like a medical device or even a drug. The consequences of a medical chatbot dysfunction could potentially have a significant negative impact, such as misdiagnosis, delayed diagnosis, inappropriate self-medication, or bad treatment adherence. Their use should not be promoted without conducting thorough investigations.

Conclusions

The data regarding the use of conversational agents in health care in general and oncology in particular are limited, which is in sharp contrast with their potential benefits for the patients and the health care system. In this phase III, blind, noninferiority, randomized controlled trial, the EORTC INFO25 scores from the chatbot were found to be noninferior to the scores of the group of physicians. Conversational agents may save patients with minor health concerns from a visit to the doctor. This could allow clinicians to spend more time to treat patients who need a consultation at the most. Consultations for symptoms that do not require an actual consultation could be avoided, potentially saving a significant amount of money and resources. However, if the quality of these computer programs is not rigorously assessed, they could be unable to actually detect the difference between minor and major symptoms, without anyone knowing. Health chatbots will need to be used by many and have access to rich datasets to increase their knowledge of medical terms, symptoms, and treatments. These systems will not replace the physicians and should be considered as a resource to enhance the efficacy of health care interventions. If chatbots are consistently shown to be effective and safe, they could be prescribed like a drug to improve patient information, monitoring, or treatment adherence. Significant hurdles still exist in the widespread application of chatbots at this time, such as compliance with the Health Insurance Portability and Accountability Act.

Acknowledgments

BC had full access to all of the data in the study and takes responsibility for the quality, integrity of the data, and the accuracy of the data analysis. The authors would like to thank Laure Guéroult Accolas, association Patients en Réseau, for her help in patient recruitment.

Authors' Contributions

JEB and BC contributed equally. JEB and BC designed the study and wrote the manuscript. BC and AP performed research. AG performed statistical analysis. The manuscript was reviewed by all the authors.

Conflicts of Interest

AG, AP, GD, BB, and PN are employed by Wefight. BC and JEB own shares of Wefight.

Multimedia Appendix 1
Questions used and adapted version of the EORTC QLQ-INFO25 questionnaire.
[DOCX File , 16 KB - jmir_v21i11e15787_app1.docx]

Multimedia Appendix 2
Detailed grading of each EORTC INFO25 item in each group.
[DOCX File , 18 KB - jmir_v21i11e15787_app2.docx]

Multimedia Appendix 3
CONSORT-EHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 2281 KB - jmir_v21i11e15787_app3.pdf]

References

1. Turing AM. Computing machinery and intelligence. Mind 1950;49:433-460. [doi: 10.1093/mind/LIX.236.433]
2. Weizenbaum J. ELIZA---a computer program for the study of natural language communication between man and machine. Commun ACM 1966;9(1):36-45. [doi: 10.1145/365153.365168]
3. Wikipedia. 2018. Dr. Sbaitso URL: https://en.wikipedia.org/w/index.php?title=Dr._Sbaitso&oldid=865309185 [accessed 2018-10-23]

4. AISB - The Society for the Study of Artificial Intelligence and Simulation of Behaviour. URL: https://aisb.org.uk/ [accessed 2018-10-23]
5. Miner AS, Milstein A, Schueller S, Hegde R, Mangurian C, Linos E. Smartphone-based conversational agents and responses to questions about mental health, interpersonal violence, and physical health. JAMA Intern Med 2016 May 1;176(5):619-625 [FREE Full text] [doi: 10.1001/jamainternmed.2016.0400] [Medline: 26974260]
6. Brouard B, Bardo P, Bonnet C, Mounier N, Vignot M, Vignot S. Mobile applications in oncology: is it possible for patients and healthcare professionals to easily identify relevant tools? Ann Med 2016 Nov;48(7):509-515. [doi: 10.1080/07853890.2016.1195010] [Medline: 27348761]
7. Laranjo L, Dunn AG, Tong HL, Kocaballi AB, Chen J, Bashir R, et al. Conversational agents in healthcare: a systematic review. J Am Med Inform Assoc 2018 Sep 1;25(9):1248-1258 [FREE Full text] [doi: 10.1093/jamia/ocy072] [Medline: 30010941]
8. Chaix B, Bibault J, Pienkowski A, Delamon G, Guillemassé A, Nectoux P, et al. When Chatbots meet patients: one-year prospective study of conversations between patients with breast cancer and a Chatbot. JMIR Cancer 2019 May 2;5(1):e12856 [FREE Full text] [doi: 10.2196/12856] [Medline: 31045505]
9. Arraras JI, Greimel E, Sezer O, Chie W, Bergenmar M, Costantini A, et al. An international validation study of the EORTC QLQ-INFO25 questionnaire: an instrument to assess the information given to cancer patients. Eur J Cancer 2010 Oct;46(15):2726-2738. [doi: 10.1016/j.ejca.2010.06.118] [Medline: 20674333]
10. Babylon Health. URL: https://www.babylonhealth.com [accessed 2018-05-25]
11. Clinical Trials. Study Buddy (an ECA Oncology Trial Advisor) for Cancer Trials URL: https://clinicaltrials.gov/ct2/show/NCT02742740 [accessed 2018-10-23]
12. Clinical Trials. Online Tailored Interventions & Relational Agents for Exercise and Sun Protection (Project RAISE) URL: https://clinicaltrials.gov/ct2/show/NCT01458002 [accessed 2018-10-23]

Abbreviations

AI: artificial intelligence
ECA: Embodied Conversational Agent
EORTC QLQ-INFO25: European Organisation for Research and Treatment of Cancer Quality of Life Group information questionnaire

Edited by G Eysenbach; submitted 07.08.19; peer-reviewed by M Hall, Y Xiao; comments to author 03.10.19; revised version received 14.10.19; accepted 23.10.19; published 27.11.19.

Please cite as:
Bibault JE, Chaix B, Guillemassé A, Cousin S, Escande A, Perrin M, Pienkowski A, Delamon G, Nectoux P, Brouard B
A Chatbot Versus Physicians to Provide Information for Patients With Breast Cancer: Blind, Randomized Controlled Noninferiority Trial
J Med Internet Res 2019;21(11):e15787
URL: http://www.jmir.org/2019/11/e15787/
doi:10.2196/15787
PMID:31774408

Original Paper

Why Health Care Professionals Belong to an Intensive Care Virtual Community: Qualitative Study

Kaye Denise Rolls[1,2,3,4*], BSc, DNurs; Margaret Mary Hansen[5*], DPhil; Debra Jackson[4,6,7*], DPhil; Doug Elliott[4*], DPhil

[1]Centre for Applied Nursing Research, University of Western Sydney, Liverpool, Australia
[2]Ingham Institute for Medical Research, Liverpool, Australia
[3]South Western Sydney Local Health District, Liverpool, Australia
[4]University of Technology Sydney, Sydney, Australia
[5]University of San Francisco, San Francisco, CA, United States
[6]Oxford Health, NHS Foundation Trust, Oxford, United Kingdom
[7]Ngangk Yira Research Centre for Aboriginal Health & Social Equity, Murdoch University, Perth, Australia
[*]all authors contributed equally

Corresponding Author:
Kaye Denise Rolls, BSc, DNurs
Centre for Applied Nursing Research
University of Western Sydney
1 Campbell Street
Liverpool, 2170
Australia
Phone: 61 2 8738 9390
Fax: 61 2 8738 9206
Email: kaye.rolls@westernsydney.edu.au

Abstract

Background: Clinical practice variation that results in poor patient outcomes remains a pressing problem for health care organizations. Some evidence suggests that a key factor may be ineffective internal and professional networks that limit knowledge exchange among health care professionals. Virtual communities have the potential to overcome professional and organizational barriers and facilitate knowledge flow.

Objective: This study aimed to explore why health care professionals belong to an exemplar virtual community, ICUConnect. The specific research objectives were to (1) understand why members join a virtual community and remain a member, (2) identify what purpose the virtual community serves in their professional lives, (3) identify how a member uses the virtual community, and (4) identify how members used the knowledge or resources shared on the virtual community.

Methods: A qualitative design, underpinned by pragmatism, was used to collect data from 3 asynchronous online focus groups and 4 key informant interviews, with participants allocated to a group based on their posting behaviors during the previous two years—between September 1, 2012, and August 31, 2014: (1) frequent (>5 times), (2) low (≤5 times), and (3) nonposters. A novel approach to focus group moderation, based on the principles of traditional focus groups, and e-moderating was developed. Thematic analysis was undertaken, applying the Diffusion of Innovation theory as the theoretical lens. NCapture (QRS International) was used to extract data from the focus groups, and NVivo was used to manage all data. A research diary and audit trail were maintained.

Results: There were 27 participants: 7 frequent posters, 13 low posters, and 7 nonposters. All participants displayed an external orientation, with the majority using other social media; however, listservs were perceived to be superior in terms of professional compatibility and complexity. The main theme was as follows: “Intensive care professionals are members of ICUConnect because by being a member of a broader community they have access to credible best-practice knowledge.” The virtual community facilitated access to all professionals caring for the critically ill and was characterized by a positive and collegial online culture. The knowledge found was credible because it was extensive and because the virtual community was moderated and sponsored by a government agency. This enabled members to benchmark and improve their unit practices and keep up to date.

Conclusions: This group of health care professionals made a strategic decision to be members of ICUConnect, as they understood that to provide up-to-date clinical practices, they needed to network with colleagues in other facilities. This demonstrated that a closed specialty-specific virtual community can create a broad heterogeneous professional network, overcoming current ineffective networks that may adversely impact knowledge exchange and creation in local practice settings. To address clinical practice variation, health care organizations can leverage low-cost social media technologies to improve interprofessional and interorganizational networks.

(J Med Internet Res 2019;21(11):e14068) doi:10.2196/14068

KEYWORDS

social media; focus groups; physician; nurse; intensive care; innovation diffusion; scholarly communication

Introduction

Background

Modern health care is delivered in complex organizations by a range of health care professions. Significant clinical practice variations may exist [1,2] in part because of ineffective internal and professional networks that limit knowledge exchange between health care professionals (HCPs) [3,4]. Virtual communities (VCs) have the potential to overcome these professional and organizational barriers [5,6], facilitating knowledge flow between HCPs and across organizations. This was the final study in a multiple-methods research program, where 3 concurrent studies examined interrelated aspects of an exemplar VC (ICUConnect): (1) the professional social network [7] (2) community participation, including knowledge exchange (manuscript under review), and (3) why HCPs join and remain members (protocol; [8]). The aim of this study was to explore *why* HCPs belong to ICUConnect.

Diffusion of Innovation

Everett Rogers [9] developed the *Diffusion of Innovation* (DOI) theory by integrating study findings from agriculture where researchers examined how individuals adopted innovations over time. Rogers then evolved the theory by undertaking studies across different countries and levels of economic and social development [9]. In health, these innovations could include new equipment, research findings, or practices. Early research focused on how the interplay between the relative characteristics of the innovation, time, and communication channels and structure of a social group affected the diffusion and adoption of that innovation over time. An innovation is an idea, practice, or object that is perceived to be new by an individual or work group, and there are 5 characteristics that influence this perception: relative advantage, complexity, compatibility, trialability, and observability [9]. Rogers found that for an innovation to diffuse across a social group, at least the first 16%, comprising innovators and the early adopters (visionaries), needed to adopt before a critical mass was reached and adoption spread to the early majority (pragmatists). The latter groups of late majority and laggards became interested in adoption when it was apparent; they were straying from group norms. A critical difference between early and late adopters is the former have greater access to new information because of the number and quality of communication channels they choose to maintain, especially outside their close social circles. For technology adoption, the gap (Moore's chasm) between the visionaries (early adopters) and pragmatists (early majority) may only be crossed when proof of the technology efficacy has been demonstrated and championed by early adopters (see Multimedia Appendix 1) [10-11].

Contemporary research has demonstrated how organizational or group factors exert a powerful influence on both individuals and the organization [12-14]. There are 7 key internal organizational factors that influence an organization's ability to develop or implement innovations, including centralization, complexity, formalization, interconnectedness, organizational slack, external orientation [9], and absorptive capacity [12,15]. Interconnectedness (connections between organizational members and units) and external orientation (organizational leaders with external networks) are both mediated by communication channels or networking internally or external to the organization [9,12-13]. Furthermore, an external orientation reflects an individual's attitude toward change, which is an independent variable when evaluating the innovativeness of an organization [9]. Individuals with communication channels outside their everyday social and professional networks will have greater access to new information because they are crossing boundaries between social groups; however, unless the source is considered credible, the veracity of information will be questioned [16]. These boundary spanning activities are vital if an organization is to have access to novel information and innovations [17]. For further description of DOI, refer to Multimedia Appendix 2.

Social Networks and Optimal Patient Outcomes

For patients to experience optimal outcomes, health care organizations must deliver clinical practices based on contemporaneous evidence. Effective identification and integration of knowledge requires organizations to balance a dense homogenous internal social network with low density diverse external social networks [12,16,18]. The prevailing vertical hierarchical structures, however, do not support the development of a cohesive, cooperative, and multidisciplinary culture necessary to address contemporary health care challenges [19]. The current reality is that significant clinical practice variability exists, leading to suboptimal patient outcomes [1]. This variability may be because of ineffective social networks [20-23] that restrict the flow of knowledge into and around a health care organization and onto individual HCPs. Some contributing factors have been identified including (1) the hierarchical organizational structure that isolates clinicians and restricts knowledge flow [4]; (2) professional boundaries between members of multidisciplinary teams that limit a shared understanding of specialty knowledge [24]; (3) workplace

socialization forcing clinicians to comply with currently accepted practices [25,26]; and (4) HCPs from a range of disciplines who prefer knowledge sources that are human, easily accessible, and perceived to be credible [27-32]. There is increasing interest in the use of social media to create these social networks as VCs that have the potential to overcome the above barriers [5], so that HCPs have sustained access to novel knowledge [33,34].

How Health Care Professionals Use Social Media to Form Virtual Communities

HCPs began using VCs in the early 1990s, although uptake of Web-based communication varies considerably across disciplines and specialties [7,35-37], and despite positive public attitudes toward what is today coined *social media*, this has not translated into significant professional use. At present, regardless of platform, the vast majority of VC members tend to not post or post infrequently; however, this is reversed when examining how often members access a VC or read posts [38]. At an individual level, members who post in an HCP VC are seeking a better understanding of the current knowledge and best practice in their particular field [39-41] or to assist fellow clinicians [36,40,42,43]. This suggests HCPs use VCs to establish virtual professional networks [13] to enhance access to colleagues and best practice knowledge. These members also develop a commitment to the VC and are motivated to post by collectivism [36,40,42,43], reciprocity [36,42,43], and where the Web-based environment is perceived to be respectful [42,43] and noncompetitive [44]. Members tend not to post when they lack time or interest, knowledge self-efficacy, confidence [41,42,45], or skills to use the platform [41,42] and when the Web-based culture or discussions are viewed unfavorably [36,41,45]. There are some data suggesting that this is influenced by individual characteristics [46], peers [46,47], and perceptions of the platform as an innovation [46]. Similar to nonhealth VCs, there is a symbiotic relationship between the online culture of a VC, members, and knowledge-sharing activities [48,49].

At present, the research base concerning the efficacy of health care VCs remains inadequate, as most of the studies concerning HCP VCs or on why or how HCPs use social media rely on Web-based observation [38], which only reveals the perspective of posters, who represent a minority of VC members. Given that regardless of professional group or industry, most VC members prefer to read rather than post [50,51], what is it that motivates HCPs to join a VC and what do they find of value that influences them to remain members? Ideally a member survey would provide data more representative of a whole community; however, prior research has struggled to obtain representative samples [52-57]. Therefore, a qualitative design was chosen because it would collect rich data from all types of members, especially the unrepresented nonposting majority. Understanding these phenomena will assist health care leaders in understanding how to develop and implement VCs to optimally leverage social media to improve knowledge diffusion and patient care.

Study Aim

The aim of this study was to explore *why* HCPs belong to ICUConnect. The related research objectives were to (1) understand why members join and remain a member, (2) identify what purpose the VC serves in in their professional lives, (3) identify how a member uses the VC, and (4) identify how members used the knowledge or resources shared on the VC.

Methods

Design

A qualitative design underpinned by pragmatism [58,59] was developed with data collected using three asynchronous and nonanonymous online focus groups and key informant interviews with participants allocated to a group based on their posting behaviors in the previous two years. The different modes of community participation by members and the symbiotic nature of the relationship between members and an individual VC [38] suggest that there is no universal VC experience. At the core of pragmatism is the acceptance of pluralism [60,61], and the value of knowledge is intrinsically dependent on the social context and values of both the research participant and scientist [61]. A range of theories have been used to develop an understanding of how or why HCPs use VCs, including the theory of planned behavior [47,62], technology acceptance model [46], and community of practice (CoP) [39,43]. The DOI theory [9] was chosen as the theoretical lens because of the need to explore the intersection between the individual member, the organization, and the innovation (ICUConnect) rather than to identify the relative importance of individual aspects. The protocol for this study has been published [8].

Ethics

A total of 2 approvals were obtained from the Human Research Ethics Committee (HREC) of the University of Technology Sydney. The first approval (HREC 2014000378) covered the online focus groups. For the online focus groups, participant confidentially was ensured by (1) a group rule, covering nondisclosure of participant names or sharing the content of posts, and participants agreed to abide when they registered for the study and (2) focus groups were convened within a secure website using a closed, password-protected discussion forum with the social media sharing function disabled. These layers were designed to ensure participant confidentiality and prevent forum posts from being searchable via the Web [63]. Informed consent for participants was included as part of the Web-based registration form. An amendment to undertake key informant interviews (HREC 2014000683) was granted because of a shortfall in recruitment for the frequent poster focus group. Participant identifying information could not be removed from the online focus groups' text; however, it was removed from transcribed interviews. All participants were given a unique identifier number to maintain a link with their original data. Confidentiality of participants was maintained by storing original data including focus group data and interviews (as MP3 files) within a university-authorized secure cloud server (Oxygen). Participant deidentification was maintained using a standardized taxonomy.

Setting

ICUConnect is a listserv, established in 2003 by a New South Wales Health state–based unit (the Intensive Care Coordination

and Monitoring Unit) to provide intensive care (IC) clinicians with online network to exchange information and improve patient care [45]. At the time of data collection, there were approximately 1600 members from all health care professions who worked at about 225 health care facilities, universities, and industry partners. Although these HCPs were from several countries, the majority were from Australia with nurses being the largest professional group [7].

Participants and Sample

A purposive stratified sampling method [64] was used to recruit the participants for the online focus groups and subsequent key informant interviews. The aim was to recruit 8 to 12 participants for each of the 3 focus groups [65,66], with focus group assignment based on Web-based participation over the preceding 2-year period (frequent: posting >5 times, low: posting ≤5 times, and nonposters: no posts). The rationale for this was to create a Web-based space where participants felt comfortable and confident that their contributions will be met in a positive and supportive environment because the other participants shared their preferred mode of participation; that is, a low or nonposter would not feel intimidated because there were no high posters who might monopolize the conversation [67]. An invitation to participate was posted on ICUConnect, with a link to the Web-based recruitment form (Google forms; Google). The recruitment form included participant information, consent, participant demographics, and a short survey covering group rules (Netiquette; refer to Multimedia Appendix 3). Once a potential participant had completed the registration and consent, their posting behavior was reviewed, and they were assigned to a focus group and notified. This review was completed by searching KDR's email archive using the potential participant's email address. Once located, the posting activities of the potential participant between September 1, 2012, and August 31, 2014, were evaluated.

Data Generation

There were 4 sources of data: (1) 3 online focus groups, (2) key informant interviews, (3) research diary, and (4) the audit trail. The first 2 components are discussed in the following section, whereas the latter 2 are discussed in the Study Methods: Strengths and Limitations section.

Moderating Focus Groups

Each focus group was conducted over 3 weeks between October and December 2014, using a closed discussion forum (IPBoard version 3 Invision, Powerboard) that was hosted on a secure jurisdictional health department website. The platform was chosen because it was accessible and usable across fixed and mobile devices. For each focus group, there were 2 weeks of active discussions, with each forum kept open for another week for any further comments. The focus groups were held in the following order: (1) low posters, (2) nonposters, and (3) frequent posters, with the low- and nonposting groups overlapping by a week.

The approach to focus group moderation was based on principles from moderating traditional focus groups [65] and facilitation of learning on the Web or electronic moderating [68] (see first table of Multimedia Appendix 4 for a priori moderating plan). KDR moderated the focus groups, and DE was a nonparticipant observer. This approach was developed a priori to maximize conditions for the development of rich data by facilitating optimal participation and interaction, and safeguarding participant confidentiality [65].

The focus group question guide was informed by *DOI* [9] and refined over time to reflect how discussions evolved (see second table of Multimedia Appendix 4 for question guide). Each question posted by the moderator formed a discrete discussion thread that explored a specific aspect of the VC including positive and negative aspects. A schedule was used with new questions posted every 2 to 3 days depending on activity. Participants were alerted to a new question by emails using a standard script with an informal and conversational tone. Elements of this standard script included the following: (1) expression of appreciation for participation, (2) reiteration that help was available if technical issues were experienced, (3) the question and any clarifying information, and (4) where applicable, summaries of previous posts that were germane to a new question.

Key Informant Interviews

A total of 4 frequent posters were purposively recruited and interviewed, between February to June 2015, to address the shortfall in the number of participants in the frequent poster focus group. A total of 3 interviews were face-to-face as participants were in metropolitan Sydney, and 1 was conducted via Skype (Skype Communications SARL, Microsoft Corporation) as the participant was located outside this area.

Data Collection

Data collected included (1) demographic data describing participant characteristics, (2) categorical data describing discussion forum participation, (3) discussion threads documenting focus group discussion, (4) transcripts of key informant interviews, and (5) field notes and research diary. Field notes recorded what the researcher experienced during data collection and included (1) both a description of and reflection on what occurred, (2) reflections on personal thoughts and feelings, and (3) any insights, judgments, and interpretations made in the field [69]. Once collected, data were stored in an NVivo file (versions 10 and 11, QRS International).

Data from the 3 online focus groups were collected using NCapture (QRS International) and imported into NVivo. The 3 face-to-face interviews were recorded on a mobile phone whereas the Skype interview was recorded using an MP3 Skype recorder (Alexander Nikiforov). These MP3 files were transcribed via a Web-based service (Transcribe Me!); following this the transcripts were anonymized and imported into NVivo for analysis. Field notes were developed concurrently with the online focus groups and during data analysis using the memo function of NVivo. An interview sheet was used to make notes during the interviews, and this was scanned and imported into NVivo.

Data Analysis

In keeping with the pragmatic realist approach, an analysis of focus group and key informant interviews was completed using

Braun and Clarke's 6-step thematic approach (this is expanded upon in Methods in Multimedia Appendix 4) [70]. DOI [9] was selected as the theoretical lens, as it aligned with both the broad problem of inadequate social networks limiting knowledge diffusion in health care, and current gaps in the literature. Member checking of early themes was undertaken during focus groups where responses could be seen to be converging.

Researcher Bias and Relationship With Participants

KDR was the long-term moderator of the VC, and DE was a member; however, the other authors were not members or associated with the VC. To manage any potential for bias during data collection and analyses and to establish a welcoming nonhierarchical atmosphere, 2 key procedures were completed. KDR withdrew from the moderator role several months before participant recruitment and completed a bracketing process [71]. This formed a part of the research diary, and these assumptions were revisited during data analyses. To mitigate for possible coercion during the focus groups, nonauthoritative language was used, and the roles of researcher and moderator (KDR) and nonparticipant observer (DE) were made explicit.

Results

This section reports study findings within the context of the DOI. The participants, including the participants as innovators, are described first, followed by ICUConnect as social media, and then presentation of the overarching theme of why HCPs belong to the VC. Participant contributions are reported verbatim except for correction of spelling and participant deidentification.

Participants

A total of 29 members enrolled for the focus groups; however, only 23 participated. Overall, there were 27 participants for this study (7 frequent posters, 13 low posters, and 7 nonposters). For the frequent poster group there were 3 from the focus groups and 4 key informant interviews (see first table of Multimedia Appendix 5). All participants had significant experience as HCPs and IC clinicians, with frequent poster participants the most experienced (see second table of Multimedia Appendix 5). Their length of professional experience suggests that all participants were digital immigrants, that is, born before 1980 [72].

Participants from the posting groups exhibited stronger external orientation or boundary spanning than nonposters, as evidenced by the frequency with which they described sharing ICUConnect discussions with colleagues inside and outside their local working environment. Low and nonposters shared a lack of knowledge self-efficacy, a preference for offline communication, and being an observer. Knowledge self-efficacy or lack of (a feeling of not having the experience or knowledge to add to a discussion) was demonstrated by the following quote:

> *I am an observer for a number of reasons...I have worked for a number of years away from the floor of the ICU...feel that I am not right up to date with the latest clinical information in the area. In my general workplace demeanor, I am reserved but definitely not a passive person.* [NUM FG2-6]

Overall, 60% (16/27) indicated they used other social media. A total of 70% (5/7) of frequent posters reported professional use of other social media compared with just over 50% for low (7/13) and nonposters (4/7). Specialty-specific VCs (discussion forums or listservs) were the most common extra social media used (26%, 7/27), followed by ResearchGate (22%, 6/27), Twitter (19%, 5/27), and podcasts or YouTube (15%, 4/27). Facebook was commonly used for personal networking only (48%, 13/27).

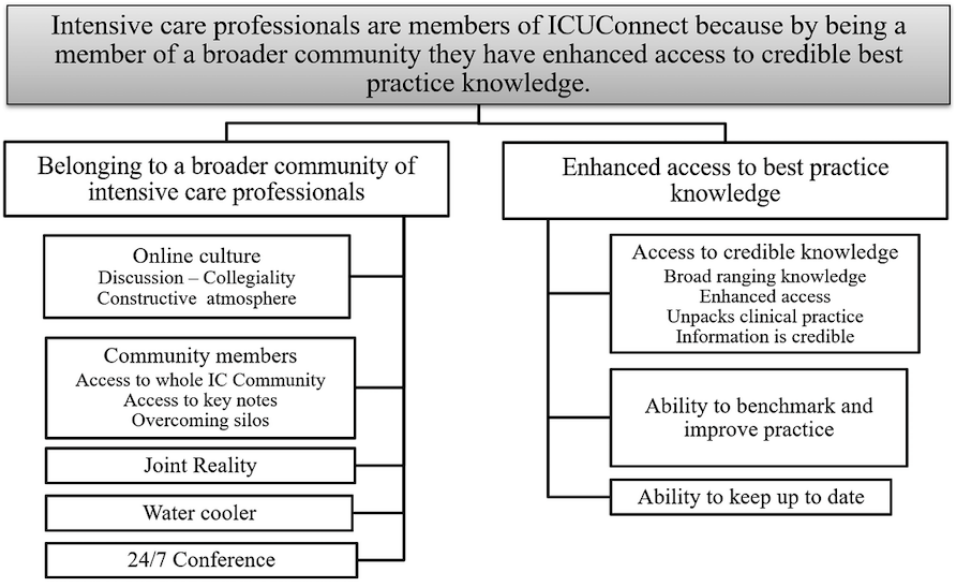
ICUConnect as Social Media

ICUConnect, an email-based listserv, was perceived by participants to be superior to other social media in terms of compatibility, complexity, and relative advantage (see Multimedia Appendix 2). Importantly, other social media were perceived as incompatible with professional values and beliefs because of the volume of information, the intrusiveness of nonprofessional information, or unprofessional language (eg, abbreviations). ICUConnect was also viewed as superior to (relative advantage) over other media because it was specific to the Australian IC context, queries were answered quickly, and the platform was perceived as being less complex to use, especially for technologically naive members.

Theme—Why We Belong

The overarching theme identified was that participants were members of ICUConnect because by belonging to a broader community of IC professionals they had enhanced access to credible best practice knowledge. A total of 2 subthemes were identified, each with elements that provided structure and context for the theme within the lens of DOI (see Figure 1): (1) *Belonging to a broader community of IC professionals* (short name: Belonging to a community) embodied the social system of ICUConnect and (2) *Enhancing access to best practice knowledge* (short name: Access to knowledge) represented how the VC facilitated innovation access for members.

Figure 1. Main theme. IC: intensive care.



Subtheme—Belonging to a Broader Intensive Care Community

This complex subtheme displayed 5 elements, as noted in Figure 1, and are discussed in further detail below. The online culture of ICUConnect was the largest subtheme element and highly valued by members regardless of posting behavior. This culture was characterized by informative discussions, collegiality, and a constructive atmosphere. Discussions were the dominant characteristic described and were viewed as being both highly and least valued by participants. When described positively, discussions were portrayed as informative and entertaining cross-disciplinary debates that provided valuable perspectives that were not available where participants worked (see Table 1, Exemplar 1). Conversely, participants also felt that discussions were limited by a lack of robust argument, nonevidence-based or ill-informed answers, that some members used discussions for self-aggrandizement, and that on occasion the content or intent of posts were misconstrued because of a lack of personal knowledge of a poster (see Table 1, Exemplar 2). The collegiality of ICUConnect was valued and was cited as a reason to join. This collegiality was exemplified by altruism (expressed by frequent posters), the willingness of members to share with colleagues, and that help was available when asked for. Importantly, this collegiality extended beyond nursing and medicine to include allied health members (see Table 1, Exemplar 3). Overall, participants felt that ICUConnect had a constructive, respectful, and informal atmosphere or tone that expedited access to knowledge, and importantly, lacked malicious interactions such as flaming or disparaging comments (see Table 1, Exemplar 4). Several participants, however, remained concerned regarding the reception of their posts (refer to third table in Multimedia Appendix 5 for more exemplars).

The second element of *Belonging to a community* was *community members*, which was characterized by 3 elements. Participants said that because ICUConnect provided *access to the whole of the IC community*, the VC made members feel a part of a broader community that simplified their networking (see Table 2, Exemplar 5). This facilitated access to IC experts (*keynotes*), which was highly valued and cited as a reason to read a post (see Table 2, Exemplar 6) and supported members in overcoming any clinical or practice silos created by local organizational structures (see Table 2, Exemplar 7). Refer to fourth table in Multimedia Appendix 5 for more exemplars.

The third element of belonging to a community was *joint reality*, where participants expressed feelings of being connected to the community, particularly when colleagues disclosed that they were experiencing similar clinical practice issues. For frequent posters this also created a sense of contributing to improving patient care on a broader scale (see Table 3, Exemplar 8). This element symbolizes a perceived homophily, that is, a sense of belonging to a like-minded group with shared values and experiences [9].

The fourth element was that ICUConnect functioned like discussions around a *watercooler* or an informal meeting place [73], where participants described using discussions to initiate conversations with work colleagues and reflect on local practices. This element was described most often by low posters but only occasionally by frequent or nonposters. As a watercooler space, ICUConnect was perceived as an extension of their local unit, with information that could be used locally or sparking and informing local discussions with new perspectives, ideas, and contemporaneous practice trends (see Table 3, Exemplar 9).

The final element was *24/7 conference*, which was a descriptor for ICUConnect because it provided immediate access to colleagues, research, and evidence; a circumstance normally limited to structured professional events such as annual conferences or seminars (see Table 3, Exemplar 10). The VC was, therefore, seen as superior or having a relative advantage over traditional professional events as it was always available and required no money or time to attend (see fifth table in Multimedia Appendix 5 for more exemplars).

Table 1. Belonging to a broader intensive care community: Element 1—online culture exemplars.

Online culture element	Exemplar
Exemplar 1: Positive discussion	*I think it's the opportunity to speak to other colleagues, be that medical or nursing, and to drill down to some of the points...at the time, you know, we all had this good debate, and I think it—I think as the debate progressed, more people came in on that discussion, and I think the wider community hopefully benefited from that. So, I think having a dialogue is of benefit.* [Equipment manager KI-2]
Exemplar 2: Negative discussion	*In terms of negatives, all I can think of (and I really had to think!) is that some posts can be misunderstood if you do not know the person posting (especially for those who are new to ICUConnect). One might say that some would be discouraged from posting, fearing a "not so favorable" reply that is FOREVER there for the whole.* [Health care manager FG3-4]
Exemplar 3: Collegiality	*Since my role has changed, I have used ICUConnect a little more to seek out advice and ideas from other areas. Much of the responses have been very positive and I have enjoyed the sharing and caring.* [Clinical nurse external FG2-5]
Exemplar 4: Constructive atmosphere	*...and I think I like principally the respectful way that people—or that they visibly deal with queries and questions and so on. And I've seen a few kind of attempts to correct direction through the years, and they've all seemed to be received well and I've agreed with them all. So I guess that it's a respectful environment that people feel really free to ask questions, sometimes over and over and over again.* [Knowledge broker KI-3]

Table 2. Belonging to a broader intensive care community: Element 2—community members.

Community members element	Exemplar
Exemplar 5: Whole of intensive care community	*ICUConnect provides me exposure to the ICU community; their thoughts; interests; discussions and topics, free of charge and easily accessible from work.* [Physiotherapist FG3-2]
Exemplar 6: Access to keynotes	*If I see a topic I may not be interested in particularly, but I see one of these people have commented, I may then read the original message and a few other comments—this gives me a quick gist of the flow of the topic, I then read the keynote response...I value the high calibre of expertise in the contributors to ICUConnect, thereby I am able to rely on information provided, or at least follow their guidance to view recommended sites to research.* [Clinical nurse—external FG3-1]
Exemplar 7: Overcoming clinical silos	*We all can get caught up in our "own world" and then we never progress, so this world allows QI to progress via discussion and research among like groups in a more timely manner.* [Clinical nurse external FG3-1]

Table 3. Belonging to a broader intensive care community: Elements 3 to 5.

Elements 3-5	Exemplar
Exemplar 8: Joint reality	*Innocent questions arise all the time and it is comforting to know that others are thinking along those same lines and asking those same questions. Some of the problems other units have made me realise I am not alone.* [Clinical nurse FG2-2]
Exemplar 9: Water cooler	*There are often interesting topics of discussion and I find that questions I have may have already been answered or ideas posed that I then take to the next level of investigation. Because I work in a small unit, with very limited resources, I find the discussions useful for formulating plans of where we should be heading. The value of this type of information sharing cannot be overstated, particularly for smaller units.* [Equipment NUM KI-1]
Exemplar 10: 24/7 conference	*Joining ICUConnect allows me to do this (gain other perspectives) from those working in the field, without having to take time out from work. I can access limited PD/study leave with virtually no funds available, so this allows me to make a contribution where appropriate on topics I can contribute to, sharing my expertise.* [Physiotherapist FG3-2]

Subtheme—Enhanced Access to Best Practice Knowledge

The second subtheme, *Enhanced access to best practice knowledge* (see Figure 1), represents how ICUConnect facilitates innovation access for members and comprised 3 elements: access to credible knowledge, being able to benchmark practice, and keeping up to date. Access to credible knowledge was a minor reason cited by participants when initially asked why they joined the VC; however, its prominence increased over the course of discussions. This element had 4 characteristics: (1) broad ranging knowledge, (2) enhanced access, (3) unpacking of clinical practice, and (4) credible information. For several members a bonus was the opportunity to access the expertise of IC leaders, referred to as *keynotes* (also previously discussed under *Community*). The most prominent characteristic of the *credible knowledge* element was access to a broad range of knowledge, including exposure to reported research that enabled participants to develop local practices and resources. When asked what specific knowledge they had obtained from the VC within the last 3 to 6 months, participants identified a comprehensive list of knowledge that included recent practice

knowledge, organizational documents, conference information, equipment, and jurisdictional newsletters. Several participants also reported that they archived discussions for later use. For some members, discussions unpacked clinical practices by introducing nuances of practice that were previously unknown or not considered (see Table 4, Exemplar 11). The second characteristic of *Access to knowledge* was that ICUConnect enhanced access to knowledge because it was a superior knowledge source (relative advantage) compared with other methods, with easy access to experts, information *simply arrived* in their email box, and that they could learn from the experience of others (see Table 4, Exemplar 12). The final characteristic was that participants considered the information credible; this was a function of access to experts or keynotes and that the VC was moderated and sponsored by a health department (see Table 4, Exemplar 13; for further exemplars refer to fifth table in Multimedia Appendix 5).

The second element of *Access to knowledge* was the *ability to compare or benchmark local practice or equipment and then improve practice*. This element was another common motivator to join ICUConnect and continued to arise over the course of focus group discussions. It was clear participants understood that it was important to gain this knowledge, including alternative perspectives, from external knowledge sources to ensure local practices reflected broadly accepted best practice. This extended beyond clinical practices to include equipment, resources, and cultural issues (see Table 5, Exemplar 11). Within this element, members sought to understand whether an innovation was worth implementation by using the experiences of fellow members or vicariously evaluating the observability and relative advantage of an innovation (see Multimedia Appendix 2). The last element of *Access to knowledge* was keeping up to date. When asked why they joined ICUConnect, many participants cited wanting to keep up to date with contemporaneous and topical knowledge (see Table 5, Exemplar 10). This was especially important for participants who did not currently work in an IC unit, as it retained a strong ongoing link to the clinical setting.

Table 4. Access to knowledge: Element 1—access to credible knowledge.

Access to credible knowledge element	Exemplar
Exemplar 11: Broad ranging knowledge	*I have used posts—I have also kept some of them...I do recall a lot of discussion on high flow oxygenation—pros & cons etc. I found this particularly interesting as we have seen a reduction in the bipap numbers and in some instances, ventilation, because of this modality.* [NUM FG1-1]
Exemplar 12: Enhanced access	*Those letters or conferences that come via the post for me tend to pile up until I get to them, but on computer, email, forums etc are readily available to me at work in down time, I do tend to get to them before I miss the application final date—or I flag them to come up so I don't forget them. So those that come in the post are often missed as I don't carry them all with me to request the day off so I can go to them, but I can request the day off immediately when looking at emails at work.* [Clinical nurse external FG3-1]
Exemplar 13: Information is credible	*As a knowledge bowerbird I value the knowledge that flows across without me having to go search for it! As I have said previously it allows me to keep a finger on the pulse and what's happening. In my current role I am on the LHD Policy and Procedure Committee and I find I call on a lot of information from ICUConnect or the ICU Best Practice Project to rebut some of the out of dated practices that people insist on - it gives me the knowledge that things have changed so I can suggest that what they are proposing is now outdated and that they need to do a literature search.***[**Knowledge broker FG1-2]

Table 5. Access to knowledge: Elements 2 to 3.

Elements 2-3	Exemplar
Exemplar 9: Benchmark and improve practice	*It is always helpful (and a relief) to know that what your unit is wanting to implement and change is on par with other practices and it is always paramount to explore why certain options are not adopted.* [Knowledge broker FG3-7]
Exemplar 2: Keeping up to date	*I saw ICUConnect as an active forum where current issues/topics would be discussed; it would be a way to keep abreast of what was going on. I think it was some time before I rustled up the courage to reply or ask for anything!***[**Knowledge broker FG2-11]

Discussion

The aim of this study was to develop an in-depth understanding of why IC HCPs were members of ICUConnect, a closed VC managed by a government agency, that is, why they join and remain a member of the VC and the purpose this plays in their professional lives and how they use the listserv and the application of knowledge sourced via discussions. The key finding was that by being a member of a broader community, they had access to credible best practice knowledge. In this context, listservs were also perceived as superior to more recent social media technology.

Listserv Technology Remains a Highly Valued and Viable Social Media Platform

ICUConnect was adopted by these participants as the listserv provided a superior way (relative advantage) of communicating with colleagues, was congruent (compatible) with professional values and beliefs, and was relatively easy to use (complexity) in comparison with other social media. These data align with evidence that HCPs prefer closed professional VCs [46] with perceived high usefulness [46,62,74] and low complexity [41,42,74]. Early research on internet technologies would suggest that a contributing factor might be that the study participants were digital immigrants and therefore perceive

newer platforms as more difficult to use [72]; however, generational differences in technology use are now under question [75,76]. In addition, it was noted that closed VCs may be a function of the need for privacy and psychological safety in a professional VC [77], with these VC types also favored by teachers [78] and health care consumers [79]. Usability (how intuitive and easy it is for members to interact within a VC) is also an integral component of ongoing community success [74,80] and was reported as an important difference between nonposters and high posters [80]. Although user needs drive individuals to experiment with social media, the perceived innovation characteristics of that media will influence final adoption decisions [77,81]. The ongoing relevance and viability of listservs can be seen its continued use by MEDLIB (a VC established in 1991) [82], the REDIRIS communities by Spanish HCPs and the health literacy discussion list [83].

ICUConnect Members Are Motivated Professionals Who Are Oriented to Change

All participants appeared to view VC membership as an integral component of professional practice, as it facilitated maintenance of a contemporaneous knowledge base. Almost two-thirds of this small group of experienced HCPs exhibited cosmopoliteness [9] because they used multiple social media channels, placing them within the early adoption groups. Although early evidence indicated limited professional use of social media by HCPs [46,56,84-88], the findings reflect more recent research where frequent posters demonstrated higher use of social media behaviors [74] and also participated in more boundary spanning activities [39]. Although there are inadequate data in this study to specifically categorize participants, their membership of ICUConnect suggest they may belong to the early adopter side of the innovator curve (see Figure 1) because they chose to communicate outside their immediate professional social network This is supported by how participants vicariously experience innovations via ICUConnect, a key characteristic of the early majority [9]. This suggests these HCPs are oriented to change, similar to a population-based study that reported a significant relationship between being open to new experiences, age, and social media use [89].

Although not all study participants were in formal leadership positions, their strategic participation in ICUConnect and use of other social media reflects an external orientation that enables them to identify innovations to incorporate into their local settings [12,90,91]. Absorption and diffusion of knowledge or innovation within an organization is the role of boundary spanners (eg, nursing unit managers or project officers) [17] and knowledge brokers (eg, nurses in education or advanced practice roles) [92]. This important boundary work contributes to organizational interconnectedness, and intellectual and social capital; reflecting necessary conditions if knowledge is to move across structural, professional, and pragmatic boundaries [19,93]. Knowledge-seeking behavior is a subjective norm shared by individuals who participant in online communities [94] and loiter in information neighborhoods [95]. This participation is likely to be strategic [96] because it is time-intensive, which has previously been identified as a barrier to posting [41,42,45]. The involvement and contributions by these individuals are not self-centered acts, rather they reflect the collegiality and altruism found in business [97] and health [38] VCs where organizational knowledge is viewed as a public good to be shared.

Intensive Care Professionals Are Members of ICUConnect Because They Belong to a Broader Community and They Have Access to Credible Best Practice Knowledge

The *belonging to a broader community* subtheme embodied the social network of ICUConnect, whereas the subtheme *enhanced access to credible knowledge* represented how the VC afforded members a superior knowledge resource compared with traditional sources. Belonging to a broader community of like-minded HCPs was an integral and highly valued component for all members. The Web-based culture was highly regarded by members because of the quality of discussions, collegiality, and informality. The social network also facilitated access to the whole of the IC community and especially to expertise from key individuals, enabling members to overcome the limitations of local clinical silos. Access to a broad range of colleagues, including experts, is an essential and highly valued aspect of both face-to-face [98] and virtual [99,100] HCP CoPs, a characteristic also common across nonhealth virtual CoPs [101]. This thematic finding adds to the current evidence, which suggests that HCPs belong to VCs to augment their access to best practice knowledge so that they remain clinically current with relevant and quality information, develop workplace resources, and benchmark practice [39,43,53,74,102,103]. Of note, this access was vital and important for all member types, not just posting members who were the main focus and participants in previous research [38]. Given that ICUConnect was in its 11th year (when the study was undertaken), these findings align with current data, which emphasize how important the relationship between a positive Web-based culture and a knowledge-sharing ethos is to the continued success of a VC [104-107].

Belonging was identified early as an integral component for a *sense of VC* [49,105,106,108], which influences how VC members develop trust and participate in Web-based knowledge-sharing activities [103-106]. Similar to Rogers' homophilly [9], *belongingness* is a contextual experience where individuals feel (1) accepted, valued, and secure within a social group; (2) connected or important to the group; and (3) their professional values align with group norms [109]. A sense of belonging creates the necessary community or relational bonds to encourage members to contribute their knowledge and expertise to the VC [44,101]. As a VC evolves, a critical mass of members see the value of sharing, where both diversity and equality are core characteristics of the online community [79]. The core elements of the overarching theme demonstrate that since it was established, ICUConnect has evolved to become a diverse multidisciplinary team social network that facilitates group affiliation by promoting a collegial professional Web-based experience.

Study Methods: Strengths and Limitations

Strengths and limitations are noted for the study. Rigor in qualitative research is a contentious space [69,110,111], with preferred terms of *trustworthiness* or *confirmability* reflecting

the accuracy and comprehensiveness in how data were collected, analyzed, and reported. To support this, an audit trail was maintained, and a clear description of the research process is provided including a thick description of participants.

Several study limitations are noted. There are 3 design elements limiting transferability to the broader population of HCPs: (1) the qualitative design using focus groups and interviews, (2) the Australian IC setting, and (3) that most participants were nurses. A quantitative design, such as a survey, may have garnered a broader representation; however, as previously noted, prior studies using surveys failed to obtain representative samples. This study instead chose to leverage the advantages of online focus groups with learnings from virtual tertiary education [68] and interviews to facilitate participation by a broad range of members, especially the previously under-researched nonposting majority. Another limitation is that the data collected may have been tilted toward positive experiences because participants were current members. A more balanced dataset may have been created by including past members, who may have different perceptions of ICUConnect; however, this was not considered feasible because past members' email addresses may have changed. This limitation may have been mitigated by specifically asking about positive and negative experiences.

A key goal of qualitative research is developing rich data and undertaking analysis that leads to findings that reflect participant experience of the phenomenon of interest. The asynchronicity of the focus groups supported moderation, researcher and participant reflexivity, and data quality and analysis. A lack of interaction in the non- and high-posting focus groups was a threat to data quality, although this was partially offset by planned strategies, which increased participation. Despite this planning, the small number of participants in the high-posting focus group [66] did reduce the contributions and interaction of this important cohort. To a limited extent, the key informant interviews may have minimized this limitation. The choice to collect data as discussion threads was a key strength and contributed significantly to study credibility and trustworthiness; namely real-time participant-controlled data collection, ensuring accurate data. Participants were also able to contribute when they wished, as discussions were not taken over by dominant talkers or experts [65], and participants had time to consider their own and previous responses, contributing to rich reflexive responses. Data analysis was enhanced by early immersion [64,70] and more time to record field notes, enabling comparison and contrasting of responses. The moderator was, therefore, able to review and reflect on responses and where appropriate, refer to participants, facilitating both member checking and early theme development.

Implications

Since the internet was established, all sectors, including business, health care, information technology, and education, have been concerned with designing VCs that optimize the user experience and achieve diverse goals such as information or resource sharing, professional development, or leveraging expertise [107,112]. The critical design elements have been established [113,114]; however, developing a bespoke platform may not ensure acceptance by a target population [115-118]. By using the DOI as the theoretical lens, this study has identified 2 antecedent factors crucial to a successful health care VC, specifically that members of the target population have an external orientation and the chosen platform is compatible with their professional norms. This implies that before implementing a VC, an organization should investigate if the intended target population have a desire to communicate with their professional colleagues using Web-based methods and which platform is acceptable.

Conclusions

The key study finding was these HCP participants were members of ICUConnect because they had access to a broader IC community, enhancing access to credible, contemporary best practice knowledge. This was a strategic move as participants understood to provide up-to-date clinical practices, they needed access to the knowledge and experience of a broad range of their colleagues. Importantly, it appeared that ICUConnect, as a closed specialty-specific VC, established a broad heterogeneous social (professional) network to overcome the current ineffective networks that adversely impact on knowledge exchange and creation in contemporary local practice settings.

Acknowledgments

KDR was supported by an Australian Government Research Training Program Scholarship. No other funds were used for the study.

Authors' Contributions

All authors participated in study design. KDR was responsible for study management and undertook the principal data analysis role. All authors reviewed this analysis and refined study findings. All authors participated in manuscript writing and approved the final version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Diffusion of Innovations—innovation adoption curve.

[PNG File , 33 KB - jmir_v21i11e14068_app1.png]

Multimedia Appendix 2
Terms used in Diffusion of Innovation (from protocol).
[PDF File (Adobe PDF File), 100 KB - jmir_v21i11e14068_app2.pdf]

Multimedia Appendix 3
Recruitment email (from protocol).
[PDF File (Adobe PDF File), 149 KB - jmir_v21i11e14068_app3.pdf]

Multimedia Appendix 4
Methods.
[PDF File (Adobe PDF File), 114 KB - jmir_v21i11e14068_app4.pdf]

Multimedia Appendix 5
Participants.
[PDF File (Adobe PDF File), 253 KB - jmir_v21i11e14068_app5.pdf]

References

1. Braithwaite J, Donaldson L. Patient safety and quality. In: Ferlie E, Montgomery K, Pederson AR, editors. The Oxford Handbook of Health Care Management. Oxford, UK: Oxford University Press; 2016:325-357.
2. Muñoz-Cañavate A, Fernández-Falero MR, Hurtado-Guapo MA. Mailing Lists on the Internet - A Collaboration Tool That is Still Alive. The Case of the REDIRIS Lists. In: Poceedings of the 16th International Conference on WWW/INTERNET. 2017 Presented at: ICWI'17; October 18-20, 2017; Vilamoura, Portugal p. 261-266.
3. McGowan BS. #SocialQI: Simple Solutions for Improving Your Healthcare. First Edition. Richmond, California: No Limit Publishing; 2012.
4. Braithwaite J. Between-group behaviour in health care: gaps, edges, boundaries, disconnections, weak ties, spaces and holes. A systematic review. BMC Health Serv Res 2010 Dec 7;10:330 [FREE Full text] [doi: 10.1186/1472-6963-10-330] [Medline: 21134295]
5. Leimeister JM, Rajagopalan B. Virtual communities: an introduction. In: Virtual Communities. London, England: ME Sharpe; 2014:3-13.
6. Currie G, White L. Inter-professional barriers and knowledge brokering in an organizational context: the case of healthcare. Organ Stud 2012 Oct 22;33(10):1333-1361. [doi: 10.1177/0170840612457617]
7. Rolls KD, Hansen M, Jackson D, Elliott D. Analysis of the social network development of a virtual community for Australian intensive care professionals. Comput Inform Nurs 2014 Nov;32(11):536-544. [doi: 10.1097/CIN.0000000000000104] [Medline: 25310223]
8. Rolls K, Hansen M, Jackson D, Elliott D. Why We Belong - Exploring membership of healthcare professionals in an intensive care virtual community via online focus groups: Rationale and protocol. JMIR Res Protoc 2016 Jun 13;5(2):e99 [FREE Full text] [doi: 10.2196/resprot.5323] [Medline: 27296929]
9. Rogers EM. Diffusion of Innovations. Fifth Edition. New York: Free Press; 2003.
10. Moore GA. Crossing the Chasm. Third Edition. New York: Harper Collins; 1991.
11. Gombault A, Allal-Chérif O, Décamps A. ICT adoption in heritage organizations: crossing the chasm. J Bus Res 2016 Nov;69(11):5135-5140. [doi: 10.1016/j.jbusres.2016.04.093]
12. Greenhalgh T, Robert G, Bate P, MacFarlane F, Kyriakidou O. Diffusion of Innovations in Health Service Organisations: A Systematic Literature Review. Oxford, UK: Blackwell Publishing; 2005.
13. Crossan MM, Apaydin M. A multi ‐ dimensional framework of organizational innovation: a systematic review of the literature. J Manag Stud 2010;47(6):1151-1191. [doi: 10.1111/j.1467-6486.2009.00880.x]
14. Fleuren M, Wiefferink K, Paulussen T. Determinants of innovation within health care organizations: literature review and Delphi study. Int J Qual Health Care 2004 Apr;16(2):107-123. [doi: 10.1093/intqhc/mzh030] [Medline: 15051705]
15. Zahra AS, George G. Absorptive capacity: a review, reconceptualization, and extension. Acad Manag Rev 2002 Apr;27(2):185-203. [doi: 10.5465/amr.2002.6587995]
16. Nieves J, Osorio J. The role of social networks in knowledge creation. Knowl Manag Res Pract 2017 Dec 19;11(1):62-77. [doi: 10.1057/kmrp.2012.28]
17. Long JC, Cunningham FC, Braithwaite J. Bridges, brokers and boundary spanners in collaborative networks: a systematic review. BMC Health Serv Res 2013 Apr 30;13:158 [FREE Full text] [doi: 10.1186/1472-6963-13-158] [Medline: 23631517]
18. Borgatti SP, Halgin DS. On network theory. Organ Sci 2011 Oct;22(5):1168-1181 Not health; so no PMID. [doi: 10.1287/orsc.1100.0641]

19. Nicolini D, Scarbrough H, Gracheva J. Communities of practice and situated learning in health care. In: Ferlie E, Montgomery K, Pedersen AR, editors. The Oxford Handbook of Health Care Management. Oxford, UK: Oxford University Press; 2016.
20. Pollack CE, Wang H, Bekelman JE, Weissman G, Epstein AJ, Liao K, et al. Physician social networks and variation in rates of complications after radical prostatectomy. Value Health 2014 Jul;17(5):611-618 [FREE Full text] [doi: 10.1016/j.jval.2014.04.011] [Medline: 25128055]
21. Hollingsworth JM, Funk RJ, Garrison SA, Owen-Smith J, Kaufman SR, Landon BE, et al. Differences between physician social networks for cardiac surgery serving communities with high versus low proportions of black residents. Med Care 2015 Feb;53(2):160-167. [doi: 10.1097/MLR.0000000000000291] [Medline: 25517071]
22. Rangachari P, Rissing P, Wagner P, Rethemeyer K, Mani C, Bystrom C, et al. A baseline study of communication networks related to evidence-based infection prevention practices in an intensive care unit. Qual Manag Health Care 2010;19(4):330-348. [doi: 10.1097/QMH.0b013e3181fa36a4] [Medline: 20924254]
23. Creswick N, Westbrook JI. Who do hospital physicians and nurses go to for advice about medications? A social network analysis and examination of prescribing error rates. J Patient Saf 2015 Sep;11(3):152-159. [doi: 10.1097/PTS.0000000000000061] [Medline: 24583953]
24. Dopson S, Locock L, Gabbay J, Ferlie E, Fitzgerald L. Evidence-based medicine and the implementation gap. Health (London) 2016 Jul 21;7(3):311-330. [doi: 10.1177/1363459303007003004]
25. Mooney M. Professional socialization: the key to survival as a newly qualified nurse. Int J Nurs Pract 2007 Apr;13(2):75-80. [doi: 10.1111/j.1440-172X.2007.00617.x] [Medline: 17394514]
26. Copnell B. The knowledgeable practice of critical care nurses: a poststructural inquiry. Int J Nurs Stud 2008 Apr;45(4):588-598. [doi: 10.1016/j.ijnurstu.2006.10.010] [Medline: 17173922]
27. Spenceley SM, O'Leary KA, Chizawsky LL, Ross AJ, Estabrooks CA. Sources of information used by nurses to inform practice: an integrative review. Int J Nurs Stud 2008 Jun;45(6):954-970. [Medline: 18677809]
28. Marshall AP, West SH, Aitken LM. Preferred information sources for clinical decision making: critical care nurses' perceptions of information accessibility and usefulness. Worldviews Evid Based Nurs 2011 Dec;8(4):224-235. [doi: 10.1111/j.1741-6787.2011.00221.x] [Medline: 21649854]
29. O'Leary DF, Mhaolrúnaigh SN. Information-seeking behaviour of nurses: where is information sought and what processes are followed? J Adv Nurs 2012 Feb;68(2):379-390. [doi: 10.1111/j.1365-2648.2011.05750.x] [Medline: 21707727]
30. Curran JA, Murphy AL, Sinclair D, McGrath P. Factors influencing rural and urban emergency clinicians' participation in an online knowledge exchange intervention. Rural Remote Health 2013;13(1):2175 [FREE Full text] [Medline: 23374031]
31. Estabrooks CA, Rutakumwa W, O'Leary KA, Profetto-McGrath J, Milner M, Levers MJ, et al. Sources of practice knowledge among nurses. Qual Health Res 2005 Apr;15(4):460-476. [doi: 10.1177/1049732304273702] [Medline: 15761093]
32. Ebenezer C. Nurses' and midwives' information behaviour: a review of literature from 1998 to 2014. New Lib World 2015 Mar 9;116(3/4):155-172. [doi: 10.1108/nlw-07-2014-0085]
33. Rau C, Neyer AK, Möslein KM. Innovation practices and their boundary-crossing mechanisms: a review and proposals for the future. Technol Anal Strateg 2012 Feb;24(2):181-217. [doi: 10.1080/09537325.2012.647647]
34. Kothari A, Hovanec N, Hastie R, Sibbald S. Lessons from the business sector for successful knowledge management in health care: a systematic review. BMC Health Serv Res 2011 Jul 25;11:173 [FREE Full text] [doi: 10.1186/1472-6963-11-173] [Medline: 21787403]
35. Kim C, Kang BS, Choi HJ, Lee YJ, Kang GH, Choi WJ, et al. Nationwide online social networking for cardiovascular care in Korea using Facebook. J Am Med Inform Assoc 2014;21(1):17-22 [FREE Full text] [doi: 10.1136/amiajnl-2012-001465] [Medline: 23703828]
36. Widemark E. Community and Learning: A Virtual Community of Practice for Nurse Practitioners. Ann Arbour, United States: Capella University; 2008.
37. Lulic I, Kovic I. Analysis of emergency physicians' Twitter accounts. Emerg Med J 2013 May;30(5):371-376. [doi: 10.1136/emermed-2012-201132] [Medline: 22634832]
38. Rolls K, Hansen M, Jackson D, Elliott D. How health care professionals use social media to create virtual communities: an integrative review. J Med Internet Res 2016 Jun 16;18(6):e166 [FREE Full text] [doi: 10.2196/jmir.5312] [Medline: 27328967]
39. Hara N, Foon Hew K. Knowledge ‐ sharing in an online community of health ‐ care professionals. Inform Tech People 2007 Aug 28;20(3):235-261. [doi: 10.1108/09593840710822859]
40. Mishori R, Levy B, Donvan B. Twitter use at a family medicine conference: analyzing #STFM13. Fam Med 2014 Sep;46(8):608-614 [FREE Full text] [doi: 10.1016/j.ajem.2012.01.020] [Medline: 25163039]
41. Tunnecliff J, Ilic D, Morgan P, Keating J, Gaida JE, Clearihan L, et al. The acceptability among health researchers and clinicians of social media to translate research evidence to clinical practice: mixed-methods survey and interview study. J Med Internet Res 2015 May 20;17(5):e119 [FREE Full text] [doi: 10.2196/jmir.4347] [Medline: 25995192]
42. Hew KF, Hara N. Knowledge sharing in online environments: a qualitative case study. J Am Soc Inf Sci 2007 Dec;58(14):2310-2324. [doi: 10.1002/asi.20698]
43. Hew KF, Hara N. An online listserv for nurse practitioners: a viable venue for continuous nursing professional development? Nurse Educ Today 2008 May;28(4):450-457. [doi: 10.1016/j.nedt.2007.07.009] [Medline: 17881096]

44. Hew KF. ProQuest | Databases, EBooks and Technology for Research. 2006. Knowledge Sharing Among Professionals in Three Online Communities URL: https://search.proquest.com/openview/9399980e8afbcde344e808111b218da3/1?pq-origsite=gscholar&cbl=18750&diss=y [accessed 2019-09-18]
45. Rolls KD, Kowal D, Elliott D, Burrell AR. Building a statewide knowledge network for clinicians in intensive care units: knowledge brokering and the NSW intensive care coordination and monitoring unit (ICCMU). Aust Crit Care 2008 Feb;21(1):29-37. [doi: 10.1016/j.aucc.2007.10.003] [Medline: 18226542]
46. McGowan BS, Wasko M, Vartabedian BS, Miller RS, Freiherr DD, Abdolrasulnia M. Understanding the factors that influence the adoption and meaningful use of social media by physicians to share medical information. J Med Internet Res 2012 Sep 24;14(5):e117 [FREE Full text] [doi: 10.2196/jmir.2138] [Medline: 23006336]
47. Archambault PM, Bilodeau A, Gagnon MP, Aubin K, Lavoie A, Lapointe J, et al. Health care professionals' beliefs about using Wiki-based reminders to promote best practices in trauma care. J Med Internet Res 2012 Apr 19;14(2):e49 [FREE Full text] [doi: 10.2196/jmir.1983] [Medline: 22515985]
48. Lin MJ, Hung S, Chen C. Fostering the determinants of knowledge sharing in professional virtual communities. Comput Hum Behav 2009 Jul;25(4):929-939. [doi: 10.1016/j.chb.2009.03.008]
49. Rotman D, Fei WP. Sense of community in virtual environments. In: Leimeister JM, Balaji R, editors. Virtual Communities. Armonk, United States: Routledge; 2014.
50. Sun N, Rau PP, Ma L. Understanding lurkers in online communities: a literature review. Comput Hum Behav 2014 Sep;38:110-117. [doi: 10.1016/j.chb.2014.05.022]
51. Edelmann N. Reviewing the definitions of 'lurkers' and some implications for online research. Cyberpsychol Behav Soc Netw 2013 Sep;16(9):645-649. [doi: 10.1089/cyber.2012.0362] [Medline: 23848960]
52. Kostagiolas P, Korfiatis N, Kourouthanasis P, Alexias G. Work-related factors influencing doctors search behaviors and trust toward medical information resources. Int J Inf Manage 2014 Apr;34(2):80-88. [doi: 10.1016/j.ijinfomgt.2013.11.009]
53. Rodriguez-Recio FJ, Sendra-Portero F. Analysis of the Spanish-speaking mailing list RADIOLOGIA. Eur J Radiol 2007 Jul;63(1):136-143. [doi: 10.1016/j.ejrad.2007.01.031] [Medline: 17344009]
54. Cook-Craig PG, Sabah Y. The role of virtual communities of practice in supporting collaborative learning among social workers. Br J Soc Work 2009 May 12;39(4):725-739. [doi: 10.1093/bjsw/bcp048]
55. Burg MA, Adorno G, Hidalgo J. An analysis of social work oncology network listserv postings on the commission of cancer's distress screening guidelines. J Psychosoc Oncol 2012;30(6):636-651. [doi: 10.1080/07347332.2012.721484] [Medline: 23101548]
56. Kukreja P, Sheehan AH, Riggins J. Use of social media by pharmacy preceptors. Am J Pharm Educ 2011 Nov 10;75(9):176 [FREE Full text] [doi: 10.5688/ajpe759176] [Medline: 22171104]
57. Usher WT. Australian health professionals' social media (web 2.0) adoption trends: early 21st century health care delivery and practice promotion. Aust J Prim Health 2012;18(1):31-41. [doi: 10.1071/PY10084] [Medline: 22394660]
58. Feilzer MY. Doing mixed methods research pragmatically: implications for the rediscovery of pragmatism as a research paradigm. J Mix Method Res 2009 Oct 22;4(1):6-16. [doi: 10.1177/1558689809349691]
59. James W. Pragmatism: a New Name for Some Old Ways of Thinking. Luxembourg, Europe: Project Gutenberg; 2013.
60. DeForge R, Shaw J. Back- and fore-grounding ontology: exploring the linkages between critical realism, pragmatism, and methodologies in health & rehabilitation sciences. Nurs Inq 2012 Mar;19(1):83-95. [doi: 10.1111/j.1440-1800.2011.00550.x] [Medline: 22212372]
61. Biddle C, Schafft KA. Axiology and anomaly in the practice of mixed methods work. J Mix Methods Res 2014 May 18;9(4):320-334. [doi: 10.1177/1558689814533157]
62. Lau AS. Hospital-based nurses' perceptions of the adoption of web 2.0 tools for knowledge sharing, learning, social interaction and the production of collective intelligence. J Med Internet Res 2011 Nov 11;13(4):e92 [FREE Full text] [doi: 10.2196/jmir.1398] [Medline: 22079851]
63. Whitehead LC. Methodological and ethical issues in internet-mediated research in the field of health: an integrated review of the literature. Soc Sci Med 2007 Aug;65(4):782-791. [doi: 10.1016/j.socscimed.2007.03.005] [Medline: 17512105]
64. Creswell JW. Qualitative Inquiry and Research Design: Choosing Among Five Approaches. Los Angeles: Sage Publications; 2014.
65. Liamputtong P. Focus Group Methodology: Principle and Practice. London: Sage Publications; 2011.
66. Hatten J, Christensen L, Liu K, Goldstone L, Thurlow M, editors. Elements of Successful Online Asynchronous Text-Based Discussions. In: Proceedings of the World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education. 2014 Presented at: ERIC'14; October 27-30, 2014; New Orleans, LA.
67. Acocella I. The focus groups in social research: advantages and disadvantages. Qual Quant 2011 Sep 3;46(4):1125-1136. [doi: 10.1007/s11135-011-9600-4]
68. Salmon G. E-moderating: The Key to Teaching and Learning Online. Third Edition. New York: Routledge; 2011.
69. Borbasi S, Jackson D. Qualitative research: the whole picture. In: Navigating the Maze of Research: Enhancing Nursing and Midwifery Practice. Third Edition. Chatswood, Australia: Mosby; 2012:123-145.
70. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]

71. Hanson DJ. The perspectives of fieldwork educators regarding level II fieldwork students. Occup Ther Health Care 2011;25(2-3):164-177. [doi: 10.3109/07380577.2011.561420] [Medline: 23899033]
72. Tapscott D. Growing Up Digital: The Rise of the Net Generation. Volume 302. New York, United States: McGraw-Hill; 2000.
73. Siu YT. eScholarship. 2015. A Virtual Water Cooler: The Ecology of an Online Community of Practice to Support Teachers' Informal Learning URL: https://escholarship.org/uc/item/9x76q5cd [accessed 2019-09-18]
74. Kay RH, Holmes JM. Exploring Virtual Communities of Practice in Healthcare Education. In: Proceedings of the World Conference on Educational Media and Technology. 2018 Presented at: EdMedia'18; June 25-29, 2018; Amsterdam, Netherlands p. 1686-1691.
75. Margaryan A, Littlejohn A, Vojt G. Are digital natives a myth or reality? University students' use of digital technologies. Comput Educ 2011 Feb;56(2):429-440. [doi: 10.1016/j.compedu.2010.09.004]
76. Helsper EJ, Eynon R. Digital natives: where is the evidence? Br Educ Res J 2010 Jun;36(3):503-520. [doi: 10.1080/01411920902989227]
77. Amichai-Hamburger Y, Gazit T, Bar-Ilan J, Perez O, Aharony N, Bronstein J, et al. Psychological factors behind the lack of participation in online discussions. Comput Hum Behav 2016 Feb;55:268-277. [doi: 10.1016/j.chb.2015.09.009]
78. Booth SE. Cultivating knowledge sharing and trust in online communities for educators. J Educ Comput Res 2012 Oct 16;47(1):1-31. [doi: 10.2190/EC.47.1.a]
79. Daneshvar H, Anderson S, Williams R, Mozaffar H. How can social media lead to co-production (co-delivery) of new services for the elderly population? A qualitative study. JMIR Hum Factors 2018 Feb 12;5(1):e5 [FREE Full text] [doi: 10.2196/humanfactors.7856] [Medline: 29434014]
80. Liao S, Chou EY. Intention to adopt knowledge through virtual communities: posters vs lurkers. Online Inform Rev 2012 Jun 15;36(3):442-461. [doi: 10.1108/14684521211241440]
81. Zolkepli IA, Kamarulzaman Y. Social media adoption: the role of media needs and innovation characteristics. Comput Hum Behav 2015 Feb;43:189-209. [doi: 10.1016/j.chb.2014.10.050]
82. Muñoz-Cañavate A, Fernández-Falero M, Hurtado-Guapo M. Information Capture and Knowledge Sharing Systems in the Field of Library and Information Science: The Case of MEDLIB-L in Medicine. In: Proceedings of the 9th International Conference on Knowledge Management and Information Sharing. 2017 Presented at: KMIS'17; November 1-3, 2017; Funchal, Portugal p. 181-188. [doi: 10.5220/0006506601810188]
83. Kurtz-Rossi S, Rikard RV, McKinney J. Learning from the field and its listserv: issues that concern health literacy practitioners. Stud Health Technol Inform 2017;240:444-463. [doi: 10.3233/978-1-61499-790-0-444] [Medline: 28972533]
84. Hughes B, Joshi I, Lemonde H, Wareham J. Junior physician's use of web 2.0 for information seeking and medical education: a qualitative study. Int J Med Inform 2009 Oct;78(10):645-655. [doi: 10.1016/j.ijmedinf.2009.04.008] [Medline: 19501017]
85. Apostolakis I, Koulierakis G, Berler A, Chryssanthou A, Varlamis I. Use of social media by healthcare professionals in Greece: an exploratory study. Int J Electron Healthc 2012;7(2):105-124. [doi: 10.1504/IJEH.2012.049873] [Medline: 23079026]
86. Deen SR, Withers A, Hellerstein DJ. Mental health practitioners' use and attitudes regarding the internet and social media. J Psychiatr Pract 2013 Nov;19(6):454-463. [doi: 10.1097/01.pra.0000438184.74359.88] [Medline: 24241499]
87. Klee D, Covey C, Zhong L. Social media beliefs and usage among family medicine residents and practicing family physicians. Fam Med 2015 Mar;47(3):222-226 [FREE Full text] [Medline: 25853534]
88. Brown J, Ryan C, Harris A. How doctors view and use social media: a national survey. J Med Internet Res 2014 Dec 2;16(12):e267 [FREE Full text] [doi: 10.2196/jmir.3589] [Medline: 25470407]
89. Correa T, Hinsley AW, de Zúñiga HG. Who interacts on the web?: the intersection of users' personality and social media use. Comput Hum Behav 2010 Mar;26(2):247-253. [doi: 10.1016/j.chb.2009.09.003]
90. Aarons GA, Hurlburt M, Horwitz SM. Advancing a conceptual model of evidence-based practice implementation in public service sectors. Adm Policy Ment Health 2011 Jan;38(1):4-23 [FREE Full text] [doi: 10.1007/s10488-010-0327-7] [Medline: 21197565]
91. Purcell R, McGrath F. The search for external knowledge. Electronic J Knowl Manag 2013;11(2):158-167 [FREE Full text]
92. Gerrish K, McDonnell A, Nolan M, Guillaume L, Kirshbaum M, Tod A. The role of advanced practice nurses in knowledge brokering as a means of promoting evidence-based practice among clinical nurses. J Adv Nurs 2011 Sep;67(9):2004-2014. [doi: 10.1111/j.1365-2648.2011.05642.x] [Medline: 21507046]
93. Vakkayil JD. Boundaries and organizations: a few considerations for research. Int J Organ Anal 2012 May 18;20(2):203-220. [doi: 10.1108/19348831211227837]
94. Lai HM, Chen CP, Chang YF. Determinants of knowledge seeking in professional virtual communities. Behav Inform Technol 2014 Jan 13;33(5):522-535. [doi: 10.1080/0144929x.2013.864709]
95. Irvine-Smith S. A series of encounters: the information behaviour of participants in a subject-based electronic discussion listA series of encounters: the information behaviour of participants in a subject-based electronic discussion list. Int J Inf Knowl Manag 2011 Nov 21;9(3):183-201. [doi: 10.1142/S0219649210002619]

96. Wendelken A, Danzinger F, Rau C, Moeslein KM. Innovation without me: why employees do (not) participate in organizational innovation communities. R D Manage 2014 Mar 3;44(2):217-236. [doi: 10.1111/radm.12042]
97. Ardichvili A. Learning and knowledge sharing in virtual communities of practice: motivators, barriers, and enablers. Adv Dev Hum Resour 2008 Jun 5;10(4):541-554. [doi: 10.1177/1523422308319536]
98. Fingrut W, Beck LA, Lo D. Oncology communities of practice: insights from a qualitative analysis. Curr Oncol 2018 Dec;25(6):378-383 [FREE Full text] [doi: 10.3747/co.25.4088] [Medline: 30607112]
99. Ting DK, Thoma B, Luckett-Gatopoulos S, Thomas A, Syed S, Bravo M, et al. CanadiEM: accessing a virtual community of practice to create a Canadian national medical education institution. AEM Educ Train 2019 Jan;3(1):86-91. [doi: 10.1002/aet2.10199] [Medline: 30680352]
100. Barnett S, Jones SC, Caton T, Iverson D, Bennett S, Robinson L. Implementing a virtual community of practice for family physician training: a mixed-methods case study. J Med Internet Res 2014 Mar 12;16(3):e83 [FREE Full text] [doi: 10.2196/jmir.3083] [Medline: 24622292]
101. Lasfer A, Vaast E. A Relational Approach on Collaborative Resource Spending in Online Communities. In: Proceedings of the International Conference on Information Systems. 2018 Presented at: ICIS;18; December 13-16, 2016; San Francisco.
102. Cervantez Thompson TL, Penprase B. RehabNurse-L: an analysis of the rehabilitation nursing LISTSERV experience. Rehabil Nurs 2004;29(2):56-61. [doi: 10.1002/j.2048-7940.2004.tb00307.x] [Medline: 15052747]
103. Schoch N, Shooshan S. Communication on a listserv for health information professionals: uses and users of MEDLIB-L. Bull Med Libr Assoc 1997 Jan;85(1):23-32 [FREE Full text] [Medline: 9028568]
104. Blanchard AL, Welbourne JL, Boughton MD. A model of online trust: the mediating role of norms and sense of virtual community. Inf Commun Soc 2011 Feb;14(1):76-106. [doi: 10.1080/13691181003739633]
105. Lin H, Fan W, Wallace L. The effects of social and technical factors on user satisfaction, sense of belonging and knowledge community usagethe effects of social and technical factors on user satisfaction, sense of belonging and knowledge community usage. Int J e-Collab 2013;9(3):13-30. [doi: 10.4018/jec.2013070102]
106. Sun Y, Fang Y, Lim KH. Understanding sustained participation in transactional virtual communities. Decis Support Syst 2012 Apr;53(1):12-22. [doi: 10.1016/j.dss.2011.10.006]
107. Ridings CM, Gefen D, Arinze B. Some antecedents and effects of trust in virtual communities. J Strateg Inf Syst 2002 Dec;11(3-4):271-295. [doi: 10.1016/s0963-8687(02)00021-5]
108. Blanchard AL. Testing a model of sense of virtual community. Comput Hum Behav 2008 Sep;24(5):2107-2123. [doi: 10.1016/j.chb.2007.10.002]
109. Levett-Jones T, Lathlean J. Belongingness: a prerequisite for nursing students' clinical learning. Nurse Educ Pract 2008 Mar;8(2):103-111. [doi: 10.1016/j.nepr.2007.04.003] [Medline: 18291327]
110. Porter S. Validity, trustworthiness and rigour: reasserting realism in qualitative research. J Adv Nurs 2007 Oct;60(1):79-86. [doi: 10.1111/j.1365-2648.2007.04360.x] [Medline: 17824942]
111. Shenton AK. Strategies for ensuring trustworthiness in qualitative research projects. Educ Inform 2004 Jul 19;22(2):63-75. [doi: 10.3233/efi-2004-22201]
112. Li H. Virtual Community Studies: A Literature Review, Synthesis and Research Agenda. In: Proceedings of the 10th Americas Conference on Information Systems. 2004 Presented at: AMCIS'04; August 6-8, 2004; New York, NY, USA URL: https://aisel.aisnet.org/amcis2004/
113. Barnett S, Jones SC, Bennett S, Iverson D, Bonney A. General practice training and virtual communities of practice - a review of the literature. BMC Fam Pract 2012 Aug 21;13:87 [FREE Full text] [doi: 10.1186/1471-2296-13-87] [Medline: 22905827]
114. Murad A, Lederman R, Bosua R, Chang S, Wark J. Design Considerations for a Virtual Community of Practice for Health Practitioners: A Learner Centred Design Approach. In: Proceedings of the 27th Australasian Conference on Information Systems. 2016 Presented at: ACIS'16; December 5-7, 2016; Wollongong, Australia.
115. Barnett S, Jones SC, Bennett S, Iverson D, Bonney A. Usefulness of a virtual community of practice and web 2.0 tools for general practice training: experiences and expectations of general practitioner registrars and supervisors. Aust J Prim Health 2013;19(4):292-296. [doi: 10.1071/PY13024] [Medline: 23823006]
116. Barnett S, Jones SC, Bennett S, Iverson D, Bonney A. Perceptions of family physician trainees and trainers regarding the usefulness of a virtual community of practice. J Med Internet Res 2013 May 10;15(5):e92 [FREE Full text] [doi: 10.2196/jmir.2555] [Medline: 23666237]
117. Curran JA, Abidi SS. Evaluation of an online discussion forum for emergency practitioners. Health Informatics J 2007 Dec;13(4):255-266. [doi: 10.1177/1460458207079834] [Medline: 18029403]
118. Brooks F, Scott P. Knowledge work in nursing and midwifery: an evaluation through computer-mediated communication. Int J Nurs Stud 2006 Jan;43(1):83-97. [doi: 10.1016/j.ijnurstu.2005.02.003] [Medline: 16326164]

Abbreviations

CoPs: communities of practice
DOI: Diffusion of Innovation

HCP: health care professional
HREC: Human Research Ethics Committee
IC: intensive care
VC: virtual community

Edited by G Eysenbach; submitted 20.03.19; peer-reviewed by A Murad, M Noordzij; comments to author 02.07.19; revised version received 26.08.19; accepted 31.08.19; published 05.11.19.

Please cite as:
Rolls KD, Hansen MM, Jackson D, Elliott D
Why Health Care Professionals Belong to an Intensive Care Virtual Community: Qualitative Study
J Med Internet Res 2019;21(11):e14068
URL: https://www.jmir.org/2019/11/e14068
doi:10.2196/14068
PMID:31687936

Original Paper

On-Demand Telemedicine as a Disruptive Health Technology: Qualitative Study Exploring Emerging Business Models and Strategies Among Early Adopter Organizations in the United States

Ryan Sterling[1], PhD; Cynthia LeRouge[2], PhD

[1]Department of Health Services, University of Washington, Seattle, WA, United States

[2]Department of Information Systems & Business Analytics, College of Business, Florida International University, Miami, FL, United States

Corresponding Author:
Cynthia LeRouge, PhD
Department of Information Systems & Business Analytics
College of Business
Florida International University
11200 SW 8th St, RB 206B
Miami, FL
United States
Phone: 1 305 348 4709
Email: clerouge@fiu.edu

Abstract

Background: On-demand telemedicine is increasingly adopted by health organizations to meet patient demand for convenient, accessible, and affordable services. Little guidance is currently available to new entrant organizations as they consider viable business models and strategies to harness the disruptive potential of on-demand telemedicine services (in particular, virtual urgent care clinics [VCCs] as a predominant and catalyst form of on-demand telemedicine).

Objective: We recognized on-demand telemedicine as a disruptive technology to explore the experiences of early adopter organizations as they launch on-demand telemedicine services and deploy business models and strategies. Focusing on VCC service lines, this study addressed the following research questions: (1) what is the emerging business model being deployed for on-demand telemedicine?; (2) what are the core components of the emerging business model for on-demand telemedicine?; and (3) what are the disruptive business strategies employed by early adopter organizations as they launch on-demand telemedicine services?

Methods: This qualitative study gathered data from 32 semistructured phone interviews with key informants from 19 VCC early adopter organizations across the United States. Interview protocols were developed based on noted dissemination and implementation science frameworks. We used the constant comparison method to transform study data into stable dimensions that revealed emerging business models, core business model components (value proposition, key resources, key processes, and profit formula), and accompanying business strategies.

Results: Early adopters are deploying business models that most closely align with a value-adding process model archetype. By and large, we found that this general model appropriately matches resources, processes, and profit formulas to support the disruptive potential of on-demand telemedicine. In total, 4 business strategy areas were discovered to particularly contribute to business model success for on-demand disruption among early adopters: fundamental disruptions to the model of care delivery; outsourcing support for on-demand services; disruptive market strategies to target potential users; and new and unexpected organizational partnerships to increase return on investment.

Conclusions: On-demand telemedicine is a potentially disruptive innovation currently in the early adopter stage of technology adoption and diffusion. On-demand telemedicine must cross into the early majority stage to truly be a positive disruption that will increase accessibility and affordability for health care consumers. Our findings provide guidance for adopter organizations as they seek to deploy viable business models and successful strategies to smooth the transition to early majority status. We present important insights for both early adopters and potential early majority organizations to better harness the disruptive potential of on-demand telemedicine.

(J Med Internet Res 2019;21(11):e14304) doi:10.2196/14304

KEYWORDS

telemedicine; disruptive technology; business model; business strategy

Introduction

Background

Health care organizations in the United States are operating in a time of high volatility [1-6]. Contributing to current pressures is the rise of consumerism in health care, driving patient demand for convenient, accessible, and affordable services. To compete and thrive, many organizations are adopting telemedicine solutions [7]. Telemedicine involves the use of medical information exchanged from one site to another via electronic communications to improve a patient's clinical health status [8]. In 2018, more than 50% of hospitals and health systems reported some form of telemedicine offering [7,9].

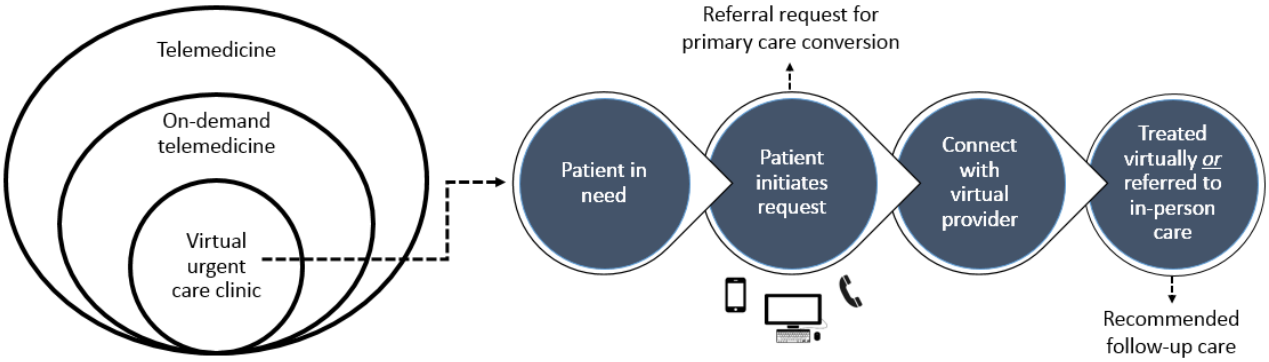
Whether telemedicine should be considered a disruptive technology is a topic of debate. Disruptive technologies are innovations that disrupt and displace established market leaders by offering products and services that are cheaper, simpler, and more convenient than what is currently available [10]. Those that assert telemedicine as a disruptive technology view it as a disruptive model of care delivery that challenges the status quo (ie, facility-based, in-person services) to create greater access and affordability in health care [11]; those in opposition view it as an innovation that improves, but ultimately sustains the performance trajectory of traditional market leaders in care delivery [12]. A holistic view of telemedicine as one health care service fuels the debate. In practice, telemedicine is not one health service offering, but actually a cadre of potential service lines, each with its own nuances in goals, workflow, stakeholders, and financing—much like in-person care.

Both practice and research may benefit from taking a closer look at forms of telemedicine that stand out as strong disrupter entrants if we want to successfully harness and leverage the potential of these service lines. It is our position that, in particular, some newer forms of telemedicine create a compelling case that they will disrupt current delivery models of medical care by offering a less-expensive, highly accessible, and more convenient alternative to many in-person options. Newcomer service lines often include offerings for *on-demand* telemedicine that are initiated by health care consumers [13]. In comparison with traditional modes of facility-based in-person care delivery, on-demand services are patient-initiated and accessible around-the-clock from any location [13]. These potential advantages may attract health organizations operating in high volatility environments, seeking ways to manage existing pressures, including the rise of consumerism. Indeed, recent research of telemedicine adoption rates and drivers indicates strong and growing interest for on-demand services that allow patients at home or on the go to reach a clinical provider for a nonemergency consult at a transparent and low-cost fee (typically US $30-50) [7].

Riding a tide of increased market growth and an uptick in adoption rates among health organizations, on-demand telemedicine may hold great promise as a disruptive technology that will bring greater accessibility and affordability to health care. However, little guidance is currently available to new entrants as they consider viable business models and strategies for on-demand services. On-demand telemedicine is in the early adopter stage of technology adoption and diffusion, with the potential trajectory of approaching early majority in the coming years [14]. This can be a precarious position for widespread assimilation of on-demand services, as the inability to bridge the innovation chasm between these stages is known to impact the success of disruptive technologies [14-16]. In general, for a disruptive technology to successfully cross into widespread assimilation, adopter organizations must understand how to navigate viable business models and strategies to expand market potential and encourage adoption among more cautious pragmatists [15,16]. Therefore, now is an opportune time to discover lessons learned from the experiences of early adopter organizations of on-demand telemedicine that are in the process of navigating these rocky waters. Few research studies in the telemedicine or disruptive technology domains provide strategy and practical guidance for those embarking on new telemedicine service lines [17,18]. Moreover, existing studies do not speak through the lens of disruptive technology to yield lessons from early adopters or detail specific forms of telemedicine [17,18].

There are many different forms of on-demand telemedicine, such as for primary care, behavioral health care, and urgent care. The virtual urgent care clinic (VCC) is a widely adopted form of the on-demand service that has received growing attention in the peer-reviewed literature [19-23]. Owing to this distinction, our study views VCC as a catalyst form of disruptive technology that can be used to examine on-demand service launch and business model deployment. VCC provides primary and urgent care services for nonemergent medical conditions that can be managed effectively by telemedicine, such as chronic bronchitis, conjunctivitis, rashes, and upper respiratory tract infections [13]. Figure 1 displays where VCC is situated in the wider context of telemedicine and reviews the general patient encounter process (see Multimedia Appendix 1 for additional information regarding the encounter process).

Figure 1. Virtual urgent care clinic encounter process.



Disruptive Technology Business Model

Although disruptive technologies have brought greater accessibility and affordability to consumers in other industries, the same cannot be widely said for the health care delivery sector [24-26]. Prior health care research suggests this failure is associated with misalignment between disruptive technologies and the need for business model innovation [25]. According to Hwang and Christensen [25]:

> *Legacy institutions of health care delivery are jumbled mixtures of multiple business models struggling to deliver value out of chaos…The health care system has trapped many disruption-enabling technologies in high-cost institutions that have conflated two and often three business models under the same roof. The situation screams for business model innovation.*

It is well documented that the success of a disruptive technology is closely tied to its business model [27-32]. The business model provides a framework for an organization to create and capture value out of the disruption [27-29]. According to Johnson et al [27], pairing disruptive technologies with the right innovative business model can lead to greater accessibility and affordability. Research indicates that business models can be generally categorized into 3 archetypes: solution shops, value-adding processes, and facilitated user networks [25,26]. Table 1 provides an overview of these leading archetypes.

To better avoid the failures encountered by other disruptive technologies in the health care delivery sector, new information is needed regarding if and where on-demand telemedicine fits into the general topology of leading business model archetypes. The current landscape of experiences among early adopter health organizations can provide us insight into emerging business models. This leads us to our first research question: *what is the emerging business model being deployed for on-demand telemedicine (specifically, in the form of VCC)?*

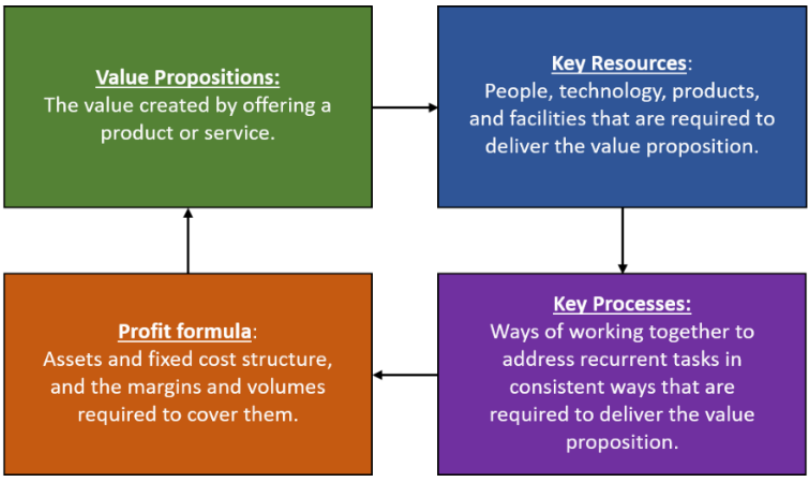
Table 1. Overview of leading business model archetypes.

Characteristics	Business model archetypes		
	Solution Shop	Value-adding process	Facilitated user network
General model description	• Used to diagnose and solve unstructured problems that are unique case to case. Value is derived from employees who diagnose causes and recommend solutions.	• Used to transform inputs into outputs of greater value. Value is derived by using standardized inputs and uniform, convenient processes to produce consistent results.	• Used for enterprises in which people exchange things with one another. Value is derived by facilitating the effective operation of the network.
Examples of model deployment	• Consulting firms • Advertising agencies • Diagnostic work performed in general hospitals	• Automobile manufacturing • Common medical procedures after definitive diagnosis	• Mutual insurance companies • eBay • Behavioral health support groups

Disruptive Technology Business Model Components

While useful, identifying a befitting type of business model does not provide the detail needed to inform strategic direction. Being leaders in the field, Johnson et al [27] understand any given business model as consisting of 4 interrelated strategic components (see Figure 2), including (1) the value proposition, or value created by offering a product or service, (2) key resources and (3) key processes that are needed to deliver the value proposition, and (4) the profit formula that defines how money is made for a deploying organization via delivery.

Figure 2. Business model framework components.



Once the 4 components coalesce into an established business model, only value propositions that fit the existing resources, processes, and profit formula can be successfully delivered [25]. Past disruptive technology research suggests these pieces must be fit together such that they are appropriately linked to an emerging disruptor for the new technology to succeed when brought to market [25]. More information is needed to specify these core components and their linkages in the context of on-demand telemedicine. To address this research gap, we propose our second research question: *what are the core components (value proposition, key resources, key processes, and profit formula) of the emerging business model for on-demand telemedicine (specifically, in the form of VCC)?*

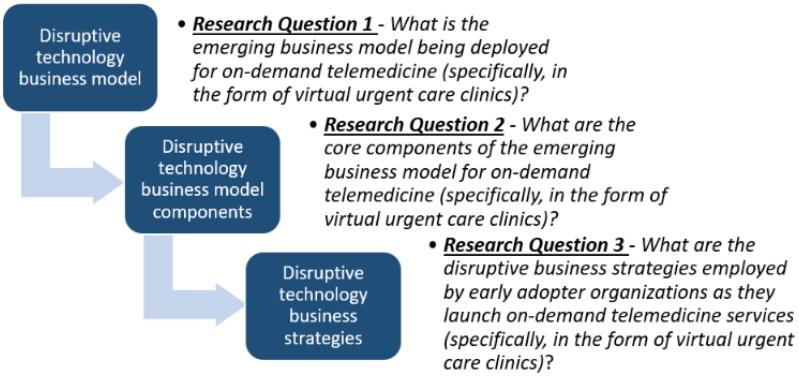
Disruptive Technology Business Strategies

While the business model describes the basic means by which an organization creates and delivers value from a disruptive technology, the business strategy is the specific method a deploying organization uses to achieve the proposed value and deal with opportunities and threats posed to the business model [28]. In the technology and innovation management field, little attention has been paid to the role of business strategies in association with emerging business models for disruptive technologies. More information is needed regarding what these disruptive strategies are and how they impact the path early adopters are taking to harness the potential of on-demand telemedicine. This leads us to our third research question: *what are the disruptive business strategies employed by early adopter organizations as they launch on-demand telemedicine services (specifically, in the form of VCC)*?

Study Objective

The objective of this qualitative study was to explore the paths that early adopters are taking to harness the disruptive potential of on-demand telemedicine, using VCC as a dominant instantiation. In doing so, we hoped to contribute to disruptive technology research by examining emerging business models and strategies being coupled with on-demand telemedicine services. We also aimed to offer practical guidance for adopter organizations as they seek to overcome some of today's leading health care challenges using disruptive telemedicine solutions. Our general research framework and specific research questions are shown in Figure 3. To our knowledge, the components of this framework have never before been studied either collectively or independently in the context of on-demand telemedicine.

Figure 3. General research framework and specific research questions.



Methods

Study Population and Data Sources

This qualitative study focuses on a study population of VCC early adopter organizations nationwide. Participants represent a range of organizational types and geographic service areas from across the United States. Table 2 provides descriptive information regarding participant organizations. In total, 5 vendor organizations are represented in our study sample (including many leading vendors among the limited number of companies currently operating in the VCC market). Among nonvendor participants, most of the organizations have contracts with vendors to provide some degree of clinical staffing and technology infrastructure to support their VCC programs.

After a 6-month national recruitment effort, we developed points of contact at 25 organizations that offer VCC services; of that total, 19 organizations (19/25, 76%) agreed to participate in our study. Convenience and purposive sampling were used to identify potential VCC adopter organizations. We targeted potential participant organizations using contact lists from the American Telemedicine Association and National Consortium of Telehealth Resource Centers. We also used Web searches to identify other organizations that may not have been listed (using keyword searches for *telemedicine*, *telehealth*, *virtual clinic*, and other related terms); Web searches resulted in identification of 2 additional participant organizations. Overall, early adopter organizations stated they were eager to participate in the confidential interview process; organizations were interested in learning from our collective, deidentified findings in publication as a means of further advancing their VCC program efforts. Among the 6 organizations that declined to participate, most of them declined because of scheduling constraints among potential key informants.

Data sources included one-hour semistructured phone interviews with key informants from participating organizations and their organization's VCC-related Web and print content. As staffing titles varied across participating organizations, organizational contacts assisted us in identifying key informants for study interviews. To recruit key informants, we targeted organizational roles related to strategy/business development, implementation, marketing, administrative operations, and clinical operations.

In total, 2 members of the research team conducted 32 phone interviews from September 2017 to December 2018. To promote an open and candid discussion, verbal and written recruiting messages emphasized confidentiality and the ability of the participants to skip questions and to go *off-the-record* with certain comments. Furthermore, at the beginning of each interview, key informants were made aware that all information collected during the interview would be completely confidential: anyone that was referred to during the interview would not be mentioned by name, nor would organizations be identified by name. All interviews were recorded (upon permission from key informants), deidentified, and transcribed before analysis. If there were any comments key informants did not wish to have recorded, the interview was postponed until all recording functions were turned off (*off-the-record*). Conversations were fluid, with few *off-the-record* requests. Failure to respond to a question was typically because of perceived lack of knowledge or factual detail related to the question; in most cases, a follow-up communication (eg, email) provided a response or a referral was made to a knowing person.

To provide breadth and depth of coverage, interview protocols were developed based on noted dissemination and implementation science frameworks that have been widely used to study the adoption of technologies in service delivery organizations, namely Damschroder Consolidated Framework for Implementation Science Research [33], Greenhalgh's framework for diffusion of innovations in service organizations [34], and Aaron's conceptual model of evidence-based practice implementation in public service sectors [35]. Collectively, these frameworks reflect a broad, sociotechnical organizational perspective that shaped our interview questions and allowed for an evidence-based exploration of business model and strategy components. Before use among key informants, experienced qualitative researchers familiar with the health information technology field and health care administrators and clinicians with a connection to telemedicine duties, such as telemedicine directors and virtual providers, reviewed the interview protocol. Minor refinements were made to the protocol as a result of this expert review (see Multimedia Appendix 2 to review our general study protocol; this general protocol was adjusted as needed to tailor interview questions and perspective to the type of organization and role of key informant).

Table 2. Characteristics of the participating virtual urgent care clinics early adopter organizations.

VCC[a] service characteristics	VCC early adopter organization type (n)			
	Health systems (n=12)	Primary care practice (n=1)	Insurer (n=1)	Vendor (n=5)
US geographic coverage				
West	4	0	1	0
Midwest	4	0	0	0
East	4	1	0	0
National	0	0	0	5
Rural/urban service area				
Urban	0	0	0	0
Rural	1	0	0	0
Urban/rural	11	1	1	5
Available VCC modalities				
Only real-time text	0	0	0	2
Only real-time phone	0	1	0	0
Only real-time video	3	0	0	1
Real-time phone and video	9	0	1	2
Vendor engagement (among nonvendors)				
Clinical staffing and other support services	10	0	1	N/A[b]
Nonclinical staffing support	2	0	0	N/A
No vendor engagement	0	1	0	N/A

[a]VCC: virtual urgent clinic care.

[b]N/A: not applicable.

Analytic Approach

We used the constant comparison method to analyze qualitative data [36,37]. Interview transcripts and supplementary Web and print content were coded independently by 1 or more research team members. Our team first deductively used noted dissemination and implementation science frameworks to develop an a priori coding schema [33-35]. Researchers met regularly during this process to iteratively discuss initial coding and refine coding categories [38]. Intercoder disagreements were resolved by consensus resolution, using an external qualitative expert to act as an auditor who makes final determinations as needed. We then carried out axial coding to inductively collapse initial coding categories into aggregate, stable dimensions that revealed emerging business models, strategic components, and accompanying business strategies [38]. Embedded in our interviewing and coding procedures, validity and reliability of study data and interpretation were assessed following Lincoln and Guba criteria for evaluating interpretive research [39,40]. Reporting of qualitative data was guided by the Consolidated Criteria for Reporting Qualitative Research [41]. We used Dedoose software for all qualitative data management and analysis [42].

Results

Overview

Our analysis revealed an emerging business model among VCC early adopters that closely aligns with the value-adding process archetype introduced in Table 1. We will first share our findings regarding the general characteristics of this emerging model and detail its core strategic components. We then describe 4 business strategies revealed from our data that are particularly indicative of the disruptive potential of VCC services.

The Emerging Business Model Deployed by Early Adopters

Identification of VCC as a value-adding process business model archetype was supported in a number of ways. First, interviewees described a general business model focused on delivering a consistent, high quality patient care experience that is quick, convenient, and highly accessibility. According to an interviewee regarding convenience, accessibility, and expediency:

> *First and foremost with [VCC], it's all about the convenience of being able to do it over your phone, your mobile phone, and on-demand. And so I've got a problem…I've got pink eye, I need to get that taken care of, I can open up my mobile phone, open up my*

> *app and I can be seen you know in less than 10 minutes.*

Regarding emphasis on consistent high-quality patient care, another interviewee commented:

> *We have defined protocols that we create based on the best literature and research out there on the appropriate way to treat patients [virtually]. We've also undertaken to hire very experienced clinicians.*

Second, indicative of the value-adding process archetype, organizations described a rule-based and uniform encounter process initiated after a VCC provider makes a definitive clinical diagnosis. Finally, with few exceptions, interviewees reported having deployed a business model dependent on service volumes to generate profit derived from the VCC encounter process. Service volumes were attributable to the VCC encounter itself and downstream from recommended follow-up care or referrals resulting from the on-demand visit:

> *So the key indicators are numbers of visits, and that includes number of visits to the website, the number of people who start the process, number of people who complete a virtual clinic visit…and then we track people who are appointed with a new primary care doctor in our system… we look at the financial return on visits that we are tracking.*

To generate volume, organizations often relied on direct-to-consumer marketing to potential users to raise awareness and drive service uptake. To accommodate the needs associated with increased service volumes, most of our participating early adopter organizations relied to some extent on vendor outsourcing to support key resource inputs for the on-demand service, such as VCC clinical staffing and/or technology infrastructure (see Table 2).

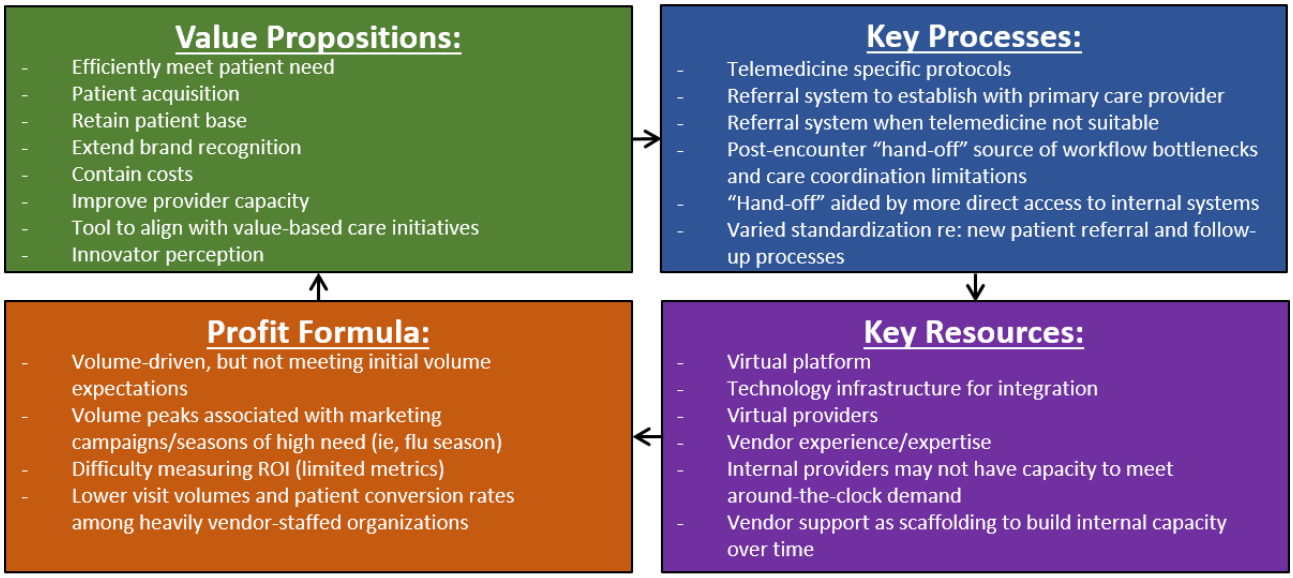
Interestingly, the collective experiences of our interviewees suggest that many early adopters are leveraging their initial investment in VCC services to explore new potential innovations in the on-demand telemedicine space that are using different business model type structures. These newly spawned innovations share elements commonly associated with the user-facilitated network business model archetype reviewed in Table 1, such as the exchange of communications and data between users, and profit generation via membership or user fees. For example, some participating organizations are cultivating VCC and other on-demand telemedicine patient user networks and technologies to manage the care of many chronic diseases. To illustrate, an interviewee described a diabetes self-management program that uses a phone-based text messaging platform to share and discuss disease management information with a wide patient community in real-time.

In addition, in alignment with the facilitated user network archetype, other participating organizations described emerging strategies to expand their membership-based service operations to increase profit generation. In such arrangements, early adopters contract with outside self-insured entities to offer VCC or other on-demand telemedicine services directly. For the self-insured entity, financial returns are achieved via improved employee health, lower employee absenteeism, and greater employee retention. For the on-demand service provider, financial returns emerge by building a larger client base.

Core Strategic Components of the Emerging Business Model

Figure 4 summarizes selected themes related to the 4 core strategic components (value proposition, key resources, key processes, and profit formula) that illuminate how VCC early adopter organizations are approaching the emerging business model we have described. We address each of these components below (a complete review of Figure 4 themes is included in Multimedia Appendix 3).

Figure 4. Summary of core strategic components of emerging business model archetype.



Value Proposition

According to interviewees, a core leading value proposition for on-demand service launch was more efficiently meeting patient need to access care. For example:

> *The value proposition for us, really comes down to better service, easier access, faster access, being mobile, you know, being able to go right where those patients are, rather than having them come to us, and really the big keyword for all of [our goals] came down to access…*

Other common value propositions included patient acquisition, retaining patient base, and extending brand recognition (often facilitated by white labeling of the VCC service by a telemedicine vendor). Regarding patient acquisition, an interviewee stated:

> *It's very expensive to acquire a new patient for health systems and so offering a convenient [virtual] urgent care and other consumer acquired services, it can be a very good way to acquire new patients and develop a new relationship with patients.*

Reducing health care costs, or cost containment, also emerged as a frequent theme. One interviewee commented:

> *…there is an incentive for the health care system to be seeing patients in this way... I think it saves [the health system] money, it saves on unnecessary costs incurred by patients being seen when they didn't have to be seen or … coming to an emergency room and utilizing resources that could better be utilized for patients who need that sort of in person service.*

Some interviewees also identified improved provider capacity as a leading value proposition:

> *For us we are having a real access issue in our small rural county. And so we were using [VCC] as a way to provide services to our community whenever we don't have provider capacity in our primary care clinic.*

While less commonly expressed by interviewees, other value propositions also included the use of VCC as a tool to support population health management (in alignment with value-based care initiatives) or to promote an innovator perception to gain competitive advantage over peer organizations. Regarding promotion of an innovator perception, an interviewee commented:

> *…health systems see the value in extending their brand, and being seen as the leader in the market of telemedicine or virtual care, it allows them to differentiate in that manner…they see this as another arm in the overall machine of trying to generate new business for the organization.*

Key Resources

According to interviewees, common key resources among early adopter organizations include the VCC virtual platform, technology middleware to link clinical and administrative systems, and virtual clinical providers to staff the on-demand service. As reviewed in Table 2, many interviewees indicated that their organizations contract with third-party vendors to source some or all of these resources, particularly clinical staffing. Vendors were seen to provide vast experience and expertise to facilitate a fast and efficient VCC launch. For example, one interviewee explained:

> *I mean if it was just putting up a video chat component that's not that difficult and anyone can do it but there is you know a lot of aspects to it, there's billing, there's claims processing, there is integration to their systems, there is doctor availability, there is managing, training…so when you come to us you kind of get that complete package plus the expertise of you know what we have been able to accomplish over the past 10 years.*

Many interviewees acknowledged that, for their organizations, pulling together key resources in-house requires extensive internal expertise regarding technology infrastructure and myriad aspects of virtual clinician staffing. While operational and clinical control was often identified as a perceived benefit, interviewees consistently indicated that it was challenging to meet around-the-clock patient demand for VCC with only their internal clinical providers. According to an interviewee:

> *Our intent is to staff it as much as possible with our employed providers. But it just doesn't make economic sense for us and we wouldn't be able to maintain a low cost point if we're having to staff [the virtual clinic] at every low utilization time, for example in the early morning. And then also we wanted to be highly accessible not just in the states where…our patients are, but have it available to those patients as they travel out of state…so we have [a] partner network [with a vendor].*

Key Processes

According to interviewees, common processes among early adopters relate primarily to the VCC encounter, including use of telemedicine specific clinical protocols and systems for primary care referral and triage to in-person services. However, interviewees reported quite varied experiences in post VCC encounter processes. Among organizations relying primarily on clinically staffing support by vendors, interviewees described a patient hand-off process between the vendor, who provides the virtual encounter, and the adopter organization, who typically handles scheduling for new referrals and follow-up to check on patient progress after the clinical encounter. As one interviewee describes this hand-off process:

> *…I mean right now it's a much more, I would say antiquated process, but the visit summary is sent to our [health information] department and then they are manually filing in that patient's chart in the media tab…So that process of getting [a patient] set up with a primary care provider is outside of the [vendor] process.*

According to interviewees, the lack of a standardized and strong hand-off process was associated with workflow bottlenecks and care coordination limitations:

> *Well ideally we would be able to get them in for a [visit] if they were hoping to have a primary care provider in our system. And so usually what ends up happening is we call them and get them on a wait list. It would be ideal if we could have more access and were able to actually pull them into our system.*

Among those organizations that do not rely primarily on clinical staffing support by vendors, most interviewees reported that postencounter processes tend to be more standardized and efficient, greatly aided by more direct access to the internal systems of adopter organizations, especially electronic medical records (EMRs), referral systems, and appointment scheduling software. As an interviewee explains:

> *I think some health care systems are adopting this model and finding it better than hiring a [vendor] simply because having it done internally, people understand the internal process, they are already utilizing the same [electronic medical record] which ends up being a huge problem with hiring a [vendor] sometimes. And so the workflow and the integration and the follow up on patient care can be a lot easier when it's done in-house rather than hiring one of these [vendors].*

Profit Formula

Overwhelmingly, interviewees described volume-driven profit generating mechanisms for VCC services, dependent on number of VCC encounters and referrals to other in-system services. However, with few exceptions, interviewees reported they are not meeting initial volume related goals:

> *I mean we're satisfied with the quality and the customer satisfaction. We are not terribly satisfied with the volume for the growth trajectory…We thought it would grow faster than it did last year.*

Volume peaks are commonly associated with VCC marketing campaigns and seasonal times of high need (ie, flu season). In general, interviewees representing organizations that rely heavily on vendor staffing typically reported lower encounter volumes and indicated less success at generating downstream volumes via patient conversion to primary care, compared with peers. As an interviewee explains:

> *[Patient] conversion is lower than what was targeted…I think we may have over projected potentially, initially on conversion.*

Review of Disruptive Business Strategies Employed by Early Adopters

Our qualitative study data revealed 4 business strategies that seem to particularly dictate the disruptive potential of VCC services, including the following: (1) fundamental disruptions to the model of care delivery; (2) outsourcing support for on-demand services; (3) disruptive market strategies to target potential users; and (4) new and unexpected organizational partnerships to increase return on investment.

Fundamental Disruptions to the Model of Care Deliver: Modern Day Twist on House Calls

Interviewees' comments regarding strategy focused on patient convenience, expediency, and appropriate level of care represent a fundamental disruption to standard models of care delivery. In fact, it can be viewed as a modern-day twist on the traditional house call. As an extension, to better facilitate the delivery of home care, many early adopters are incorporating home-based diagnostic testing and smartphone-based tools and peripheral devices to extend the capabilities and conveniences of VCC services:

> *I think we'll continue to see services evolve more and more to bring the online experience into a connected experience in the home…There are many devices available that you can attach to your Smartphone that would enable the provider to look in an ear or to listen to your heart or to listen to your lungs…and devices for home lab testing. So yeah it's something that we are keeping an eye on and then also thinking of how we can best utilize those to extend our services…[it's] definitely something we are watching.*

Regarding displacing traditional models, our data revealed a priority on right fitting care via the VCC care delivery model. One participating organization described placing VCC kiosks near emergency department waiting rooms to help triage patients to appropriate care settings based on medical need and patient choice:

> *We are looking at putting in a ER kiosk for virtual visits in one of our rural hospitals…that leadership team is wanting to have an option for those that really don't need an ER visit that are using it more for primary care, to give them an option of a virtual visit…if it's determined that really that patient does not need an ER visit, then they will be given options of seeing an ER physician, a same day appointment with the primary care doctor, urgent care option, or a virtual visit…and they'll be given the cost.*

Outsourcing Support for On-Demand Services

As reported by many interviewees, early adopter organizations often outsource to third party vendors to launch, operate, and maintain their VCC services. According to our findings, outsourcing of clinical services is a relatively new and disruptive practice for adopting organizations. Early adopters reported varied and often flexible contracting relationships with vendors, particularly around support for clinical staffing. Although some limitations around the use of vendor services were noted, specifically lack of direct access to the internal EMR and billing systems of adopter organizations, vendor experience and expertise was largely considered a useful and agile resource for early adopters to expediently launch VCC services and to provide virtual clinical provider capacity for their VCC programs.

However, a complete dependency on external virtual clinical providers to staff the service line was not a permanent strategy. Many interviewees reported outsourcing strategies that utilized varying degrees of vendor support to provide important virtual

provider scaffolding and increasingly bring the VCC service in-house as internal capacity improves and patient base expands. According to one interviewee:

> *While we could build it in house, our IT currently doesn't have a skill set to be able to sport something of this magnitude…Now that being said, I know we are currently in discussions and are working on a plan, that hopefully within the next six to 12 months, that will start to combine [vendor] providers with our own.*

Disruptive Market Strategies to Target Potential Users

Owing to the patient-initiated nature of VCC and other on-demand telemedicine services, direct marketing to potential users emerged as a central and disruptive theme in the business strategies described by early adopters. Collectively, interviewees reported that VCC marketing strategies were largely new and uncharted terrain for their staff, distinct from the marketing needs for facility-based care delivery of in-person services:

> *Getting the name out there that was something we've never really had to do before. Because usually it's just our name since health care is usually a new office, and [patients] already know what that health care is, [they] already know what an office does we don't have to really educate or re-educate. [However, this was] a brand new product, brand new service, we had to get our name out there and educate [potential users] on what the product was and how it worked.*

Interviewees overwhelmingly commented on the importance of *direct-to-consumer* marketing strategies to raise service awareness among potential users and ultimately drive service utilization and uptake. According to an interviewee:

> *We talk to clients about marketing all the time! Keeping that in their ear because, when it comes down to the bottom line, that's what really drives utilization…Always, on our agenda every week we ask, what's your marketing, what discussions are you having, this did not work so what can we do differently to make sure it works.*

Early adopters reported the use of varied marketing strategies, both traditional (eg, billboards and radio) and digital (eg, search engine optimization and websites). Interviewees reported marketing success when they prioritized funding and staffing for marketing efforts during initial VCC implementation as well as on an ongoing basis and utilized diverse marketing strategies, both traditional and digital. We further identified the value of marketing campaigns to specific seasons (eg, flu season) or opportunities of need (eg, part of information packets sent to new and relocated employees).

New and Unexpected Organizational Partnerships

To increase opportunities for return on investment from VCC service launch, and to drive profit generation, many early adopters described new, and often surprising, partnerships with organizations outside of traditional health care delivery sector circles. For example, as discussed above, some interviewees commented on future plans to expand membership operations by partnering and contracting with self-insured organizations to offer VCC services directly and at a fee. According to an interviewee:

> *[Health systems are looking to] expand to a member program or a direct to employer program…there's a huge opportunity there where a health system can go out and sell their brand name to these other organizations within the area.*

As another example of the unique partnerships undertaken by early adopters, an interviewee discussed contracting with a nationwide hotel chain to offer VCC services to guests and employees. These new partnership strategies are innovative for the health care delivery sector and appear to be supporting many early adopters in their attempts to leverage value from their VCC services.

Discussion

Principal Findings

This qualitative study used the dominant instantiation of VCC to explore the paths that early adopter organizations are taking to harness the disruptive potential of on-demand telemedicine. In the coming years, this arguably disruptive form of telemedicine will seek to attract an early majority category of adopters. In turn, our findings contribute to the literature by providing insight for researchers and organizations considering launch or expansion of on-demand services to leverage what early adopter organizations have learned along the way regarding business model deployment. We also offer practical lessons learned regarding key strategy choices for adopter organizations as they launch on-demand services and encounter hurdles to value capture and delivery via deployed business models.

Insights Into the Emerging Business Model for On-Demand Telemedicine

Health organizations have traditionally faced many struggles in aligning disruptive technologies with innovative business models [24-26]. To better understand whether organizations launching disruptive on-demand telemedicine services will meet a similar fate, this study explored emerging business models in the context of VCC early adopter organizations. With few exceptions, our study data suggest that current VCC early adopters are deploying value-adding process models that appear to appropriately match resources, processes, and profit formula to support value propositions for on-demand telemedicine.

By disentangling the reports from our interviewees regarding various business model archetypes, we were able to see a visionary progression of innovation among early adopters. Our findings demonstrate that business model archetypes and model components may evolve as organizations encounter challenges and opportunities related to VCC as a disruptive technology. In our study, we see many VCC early adopters that originally deployed a value-adding process model archetype beginning to transition to the use of a user-facilitated network model to better capture market share. To continue riding the wave of disruptive innovation and expansion spawned by on-demand telemedicine,

early adopters are not staying stagnant: they are continuing to evolve their business models and recalibrate their core model components and strategies as new challenges and opportunities arise. Future research should pay particular attention to the deployment of user-facilitated networks, as many of the early adopters participating in our study indicated increasing use of this archetype as they explore new potential on-demand telemedicine innovations within their organizations.

Strategic Direction: Strategy Helps to Transform the Business Model Into Action

We identified 4 strategy areas that seem to particularly dictate the disruptive potential of VCC services, including innovations in care delivery, outsourcing support, marketing strategies, and unique organizational partnerships. Below we review lessons learned for each of these strategy areas to help guide future practice for VCC and other forms of on-demand telemedicine.

Innovations in Care Delivery

Through much of the early 1900s, roughly half of all clinical visits involved a doctor coming into a patient's home [43]. As health care systems grew larger, more specialized, and complex over the next century, the practice of the traditional house call became nearly nonexistent; facility-based, more expensive and often time-consuming models of care delivery, such as the physician office visit and emergency department, moved in to take its place [43]. On-demand telemedicine represents a fundamental change in the model of care delivery for patients—a modern-day re-envisioning of the traditional house call. Presently, VCC and other on-demand telemedicine services are pointing back to home care as a low-cost way to reduce time constraints, improve convenience and accessibility, and engage in shared decision making with patients to *right fit* care for common nonemergent conditions.

This new delivery model presents clear gains in convenience and accessibility for the treatment of many common, nonemergent medical conditions. However, when follow-up services are required to check on patient progress or to schedule patient appointments after the on-demand visit, our findings identified workflow bottlenecks and care coordination limitations within the postencounter process for many early adopter organizations. This may indicate a struggle to integrate home-based services into the larger continuum of care when patient contact and care coordination services are needed beyond the initial virtual visit.

There is limited guidance in the research literature regarding this integration process to inform decision making among adopting health organizations. However, lessons learned from our participating early adopters suggest that clinical integration of virtual visits into patient EMRs and other electronic systems to help track patient history and facilitate care coordination needs may be an important step to strengthen postencounter processes and the new care delivery model as a whole. Recently proposed policy by the Centers for Medicaid and Medicare Services—that will give patients access to their own downloadable health data [44]—may have implications that will break down barriers to the exchange of EMR data in the near future. The proposed initiative will potentially circumvent the EMR to empower health care consumers to share their health data with whomever they wish, including virtual providers.

Outsourcing Support

Among early adopters, outsourcing to third-party telemedicine vendors emerged as a key strategy to increase speed to market, gain access to technical infrastructure without taxing internal resources, and extend clinical staffing coverage for the on-demand service. Although interviewees described a variety of outsourcing contract arrangements, those that balanced internal resources with important scaffolding support from vendors appeared best suited to meet proposed value propositions. Outsourcing clinical services is still a relatively new concept to the health care delivery sector, and as such, there is limited guidance to inform future outsourcing decisions from telemedicine and health care sources. However, findings from the wider literature may prove instructive in the context of on-demand telemedicine [45-59]. Evidence-based guidance from the general outsourcing literature suggests adopter organizations should consider outsourcing a service in the context of low internal resources (particularly human resources) [48,49], the desire to increase flexibility regarding resources, operations, and other strategic elements [50], high internal costs (relative to expected costs of outsourcing) [51,52], and if other competitors are already outsourcing a given service [53]. In contrast, evidence suggests organizations should shy away from outsourcing a service in the context of high levels of market uncertainty [54], heavy integration of the service into internal systems [55,56], high level of service complexity [57], and if the service is considered a core competency to the service line [58,59]. We call VCC organizations and ensuing research to consider this evidence-based outsourcing guidance from other domains in exploring future strategies.

Marketing Strategies

Recent health care trends indicate overall telemedicine use is growing fast among patients but remains low overall [60]. These trends were echoed in what we heard from early adopters in our study, where most of the interviewees indicated that though their VCC service volumes were increasing, they were not meeting initial projections. Low utilization does not seem to be associated with usability issues [61,62] nor dissatisfaction [63], which have been identified as some of the more common barriers to technology adoption and use. In fact, many of our participants used patient satisfaction surveys as a means to measure satisfaction as an outcome and reported that patients that used VCC services were very satisfied. Upon investigating the few reports of dissatisfaction, the most often indicated underlying cause was the patient not receiving a prescription for antibiotics when they wanted one.

Instead, with few exceptions, early adopters connected their lower than expected VCC volumes to challenges around raising awareness for the service among potential users; to address awareness, interviewees often commented on the importance of *direct-to-consumer* marketing efforts. The importance of raising awareness of a new innovation is not new to disruptive technology research: awareness and knowledge generation is considered the first step in deciding whether to use a new innovation [14]. Not addressing awareness issues can impede

adoption of consumer health technologies [64]. Increased awareness is often driven by the intersection of need recognition and marketing communications [14].

However, as was recognized in our study data, VCC marketing is largely new and uncharted terrain for early adopter organizations; according to an interviewee:

> *Getting the [VCC] name out there that was something we've never really had to do before. Because usually it's just our [organization] name since health care is usually a new office, and [patients] already know what that health care is…*

VCC marketing efforts seem to have a 3-fold purpose: (1) to provide the health consumer with understanding about the availability of VCC; (2) educate the health consumer about the medical situations when VCC is a good option; and (3) *sell* the health organization as this is where a strong link needs to be created for the health consumer to turn to the health system's VCC offering among other options. Regarding education, as with some other early innovations (eg, LinkedIn), potential adopters may not understand all of the uses and potential of VCC.

Marketing in the form of health system *branding* is still relatively new, and marketing direct-to-consumer services like VCC are even newer. In cases of one-time or episodic care similar to VCC (where the patient may not always interact with the same provider), research suggests that the presence of tight bonds between patients and a sponsoring organization, or even organizational representatives, is a key facilitating factor for successful telemedicine service interactions [61]. This finding has important potential implications for organizations as they market their VCC services. First, organizations should consider directing their marketing efforts not only toward potential virtual patients but also organizational representatives (ie, primary care providers, other staff) who may share their existing close bonds with their patients and can function as pseudo brand ambassadors to raise awareness of VCC services. We also learned in our conversations with interviewees of some limited activity in this area, particularly in regard to adopter organizations asking physicians to post VCC advertisements in their offices. Second, it indicates that as health organizations continue to expand and strengthening their *health organization branding*, they should leverage their organizational brand in their marketing efforts to raise awareness for VCC; they should consider marketing VCC not as a separate product, but instead as an available service offered by an organization that patients already know and trust to manage their medical care. Building this type of patient-organization connection is still relatively new and evolving, as patients are generally more welded to individual providers rather than to health organizations. Adopting provider organizations, such as health systems, may have an advantage in leveraging patient-organization relationships to raise VCC awareness because of their potential role as a regular source of in-person care for patients and as a well-known health care institution in local communities. We see that some early adopters are already engaging in this activity by working with vendors to white-label their VCC services so that they may present the service with strong health system branding.

However, early adopter organizations should also recognize important externals factors that may present challenges to ongoing marketing efforts to raise VCC awareness and drive utilization; namely, limited telemedicine reimbursement that may prevent penetration to certain patient markets (eg, Medicare patients), and provider credentialing and other regulations that may prevent organizations from providing services across state lines [65,66]. Although recent policy changes have reduced these limitations [67], policy barriers are not completely eliminated, and those still challenge the capabilities of health organizations adopting VCC to expand virtual service offerings and grow their patient volume.

Although, in our study, we identified a number of strategies that led to greater marketing success among VCC early adopters to drive uptake (eg, using both traditional and digital strategies), there is little additional evidence-based guidance to inform future strategic decision making in the health care marketing literature, creating an opportunity for future work. Future research efforts may be informed by research exploring factors to help organizations design, manage, and market service delivery interactions for medical video conferencing, a different form of telemedicine [68].

Unique Partnerships

According to interviewees, early adopter organizations are particularly motivated to explore innovative relationships with external entities to increase the opportunity for return on investment and profit generation related to on-demand telemedicine services. Reviewed above, a prominent example of this involves early adopter health systems contracting with self-insured organizations to offer VCC services directly. Examining other emerging and unexpected partnerships between health care and business entities, such as the recent formation of a health care company between Amazon, Berkshire Hathaway, and JPMorgan, may help to shed some light on how these innovative organizational relationships will influence the direction of VCC and other telemedicine services in the future. With the goal of improving health care services and cutting costs for more than 1.1 million employees, the Amazon partnership is predicted to disrupt the health care marketplace by using technology solutions to develop innovative treatments and modernize delivery system processes [69]. Similarly, new partnership arrangements related to VCC and other on-demand telemedicine solutions also have the potential to disrupt health care. It remains to be seen how these new organizational relationships may impact the use of various business model archetypes and strategies for new technologies in health care.

Study Limitations

Our focus on a narrow study population of VCC early adopter organizations may limit the generalizability of our study findings. As a result, some findings may not be applicable to other forms of on-demand telemedicine, such as behavioral health. In addition, we did not study nonadopters or organizations with failed VCC adoption experiences; learning about the experiences and challenges faced by these

organizations would have provided additionally meaningful insights to address our research objective. Our use of a convenience and purposive sampling approach may also present limitations to study generalizability. Although we targeted organizations in different geographic areas and of varying size and type, it is possible that the perspectives of some VCC early adopters are not represented in our study dataset. It is also possible that given time constraints, lack of knowledge, or hesitancy to discuss business information, key informants may not have shared some details of potential interest to researchers. However, key informants were generally very open and forthcoming during study interviews, thus reducing concerns that important themes may not have been revealed. They were eager to share and indicated they were motivated to learn from our findings as a means to further advance their VCC program efforts. Finally, we specifically targeted *early adopters*, representing only a minority of potential adopters along Rogers diffusion of innovation curve [14]. However, the purpose of our study was to offer guidance to new organization entrants as they consider viable business models and strategies for on-demand telemedicine, necessitating an exclusive focus on early adopters.

Conclusions

Current trends suggest health organizations will increasingly use on-demand telemedicine as a means to meet patient demand for convenient, accessible, and affordable services, and to address other leading health care challenges. Here we presented on-demand telemedicine as a potentially disruptive innovation in the early adopter stage of technology adoption and diffusion. For the research community, we contributed a new level of contextualization to disruptive innovation research targeted to the health information technology space. For early adopters, the insights we have shared can help organizations navigate evolving opportunities and address challenges to leverage their position of early entry. However, to truly be a positive disruption that will increase accessibility and affordability for health care consumers, on-demand telemedicine must cross into the early majority stage of widespread assimilation. For potential early majority organizations that are considering launch of on-demand services, insights from this study provide an opportunity to leverage what early adopters have already learned along the way to mitigate unknowns and risks as they deploy innovative business models and make strategy choices to harness the disruptive potential of on-demand telemedicine.

Acknowledgments

This research was funded in part by a grant from the Centers for Health Organization Transformation.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Overview of virtual urgent care clinic patient encounter process.
[DOCX File , 14 KB - jmir_v21i11e14304_app1.docx]

Multimedia Appendix 2
General interview protocol.
[DOCX File , 22 KB - jmir_v21i11e14304_app2.docx]

Multimedia Appendix 3
Summary of core strategic components of emerging business model archetype.
[DOCX File , 19 KB - jmir_v21i11e14304_app3.docx]

References

1. Centers for Medicare & Medicaid Services. 2018. National Health Expenditure Data URL: https://www.cms.gov/research-statistics-data-and-systems/statistics-trends-and-reports/nationalhealthexpenddata/nationalhealthaccountshistorical.html [accessed 2019-09-25] [WebCite Cache ID 77R88xiz3]
2. Anderson GF, Reinhardt UE, Hussey PS, Petrosyan V. It's the prices, stupid: why the United States is so different from other countries. Health Aff (Millwood) 2003;22(3):89-105. [doi: 10.1377/hlthaff.22.3.89] [Medline: 12757275]
3. Papanicolas I, Woskie LR, Jha AK. Health care spending in the United States and other high-income countries. J Am Med Assoc 2018 Mar 13;319(10):1024-1039. [doi: 10.1001/jama.2018.1150] [Medline: 29536101]
4. Schwartz MD. Health care reform and the primary care workforce bottleneck. J Gen Intern Med 2012 Apr;27(4):469-472 [FREE Full text] [doi: 10.1007/s11606-011-1921-4] [Medline: 22042605]
5. Petterson SM, Liaw WR, Phillips RL, Rabin DL, Meyers DS, Bazemore AW. Projecting US primary care physician workforce needs: 2010-2025. Ann Fam Med 2012;10(6):503-509 [FREE Full text] [doi: 10.1370/afm.1431] [Medline: 23149526]
6. Mehrotra A. The convenience revolution for treatment of low-acuity conditions. J Am Med Assoc 2013 Jul 3;310(1):35-36 [FREE Full text] [doi: 10.1001/jama.2013.6825] [Medline: 23821082]

7. Sage Growth. Defining Telemedicine's Role: The View From the C-suite URL: http://go.sage-growth.com/defining-telemedicine [accessed 2019-04-06] [WebCite Cache ID 77R8UwFz0]
8. American Telemedicine Association. 2018. Telemedicine Glossary URL: http://thesource.americantelemed.org/resources/telemedicine-glossary [accessed 2019-09-25]
9. Reach Health. U.S. Telemedicine industry benchmark survey. 2018. URL: https://reachhealth.com/resources/telemedicine-industry-survey/ [accessed 2019-09-25]
10. Christensen CM. The Innovator's Dilemma: When New Technologies Cause Great Firms To Fail. Boston: Harvard Business Review Press; 2013.
11. Schwamm LH. Telehealth: seven strategies to successfully implement disruptive technology and transform health care. Health Aff (Millwood) 2014 Feb;33(2):200-206. [doi: 10.1377/hlthaff.2013.1021] [Medline: 24493761]
12. Bashshur RL, Shannon G, Krupinski EA, Grigsby J. Sustaining and realizing the promise of telemedicine. Telemed J E Health 2013 May;19(5):339-345 [FREE Full text] [doi: 10.1089/tmj.2012.0282] [Medline: 23289907]
13. Gough F, Budhrani S, Cohn E, Dappen A, Leenknecht C, Lewis B, et al. ATA practice guidelines for live, on-demand primary and urgent care. Telemed J E Health 2015 Mar;21(3):233-241. [doi: 10.1089/tmj.2015.0008] [Medline: 25658882]
14. Rogers EM. Diffusion Of Innovations. New York: Free Press; 2003.
15. Moore G. Crossing The Chasm: Marketing And Selling High-tech Products To Mainstream Customers. New York: HarperCollins; 2014.
16. Ram S, Jung H. Innovativeness in product usage: a comparison of early adopters and early majority. Psychol Mark 1994 Jan;11(1):57-67. [doi: 10.1002/mar.4220110107]
17. Peters C, Blohm I, Leimeister JM. Anatomy of successful business models for complex services: insights from the telemedicine field. J Manag Inform Syst 2015;32(3):75-104. [doi: 10.1080/07421222.2015.1095034]
18. Chen S, Cheng A, Mehta K. A review of telemedicine business models. Telemed J E Health 2013 Apr;19(4):287-297. [doi: 10.1089/tmj.2012.0172] [Medline: 23540278]
19. Ashwood JS, Mehrotra A, Cowling D, Uscher-Pines L. Direct-to-consumer telehealth may increase access to care but does not decrease spending. Health Aff (Millwood) 2017 Mar 1;36(3):485-491. [doi: 10.1377/hlthaff.2016.1130] [Medline: 28264950]
20. Uscher-Pines L, Mehrotra A. Analysis of Teladoc use seems to indicate expanded access to care for patients without prior connection to a provider. Health Aff (Millwood) 2014 Feb;33(2):258-264. [doi: 10.1377/hlthaff.2013.0989] [Medline: 24493769]
21. Uscher-Pines L, Mulcahy A, Cowling D, Hunter G, Burns R, Mehrotra A. Antibiotic prescribing for acute respiratory infections in direct-to-consumer telemedicine visits. JAMA Intern Med 2015 Jul;175(7):1234-1235. [doi: 10.1001/jamainternmed.2015.2024] [Medline: 26011763]
22. Gordon AS, Adamson WC, DeVries AR. Virtual visits for acute, nonurgent care: a claims analysis of episode-level utilization. J Med Internet Res 2017 Feb 17;19(2):e35 [FREE Full text] [doi: 10.2196/jmir.6783] [Medline: 28213342]
23. Uscher-Pines L, Mulcahy A, Cowling D, Hunter G, Burns R, Mehrotra A. Access and quality of care in direct-to-consumer telemedicine. Telemed J E Health 2016 Apr;22(4):282-287 [FREE Full text] [doi: 10.1089/tmj.2015.0079] [Medline: 26488151]
24. Herzlinger RE. Why innovation in health care is so hard. Harv Bus Rev 2006 May;84(5):58-66. [Medline: 16649698]
25. Hwang J, Christensen CM. Disruptive innovation in health care delivery: a framework for business-model innovation. Health Aff (Millwood) 2008;27(5):1329-1335. [doi: 10.1377/hlthaff.27.5.1329] [Medline: 18780919]
26. Christensen C, Grossman J, Hwang JH. The Innovator's Prescription: A Disruptive Solution For Health Care. New York: Mcgraw-hill Education; 2009.
27. Johnson M, Christensen C, Kagermann H. Reinventing your business model. Harv Bus Rev 2008;86(12):57-68 [FREE Full text]
28. Chesbrough H, Rosenbloom R. The role of the business model in capturing value from innovation: evidence from Xerox Corporation's technology spin-off companies. Ind Corp Change 2002;11(3):529-555. [doi: 10.1093/icc/11.3.529]
29. Amit R, Zott C. Value creation in E-business. Strat Mgmt J 2001;22(6-7):493-520. [doi: 10.1002/smj.187]
30. DaSilva CM, Trkman P, Desouza K, Lindič J. Disruptive technologies: a business model perspective on cloud computing. Technol Anal Strateg 2013;25(10):1161-1173. [doi: 10.1080/09537325.2013.843661]
31. Schneider S, Spieth P. Business model innovation: towards an integrated future research agenda. Int J Innov Mgt 2013 Mar 18;17(01):1340001. [doi: 10.1142/S136391961340001X]
32. Zott C, Amit R, Massa L. The business model: recent developments and future research. J Manag 2011 May 2;37(4):1019-1042. [doi: 10.1177/0149206311406265]
33. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implement Sci 2009 Aug 7;4:50 [FREE Full text] [doi: 10.1186/1748-5908-4-50] [Medline: 19664226]
34. Greenhalgh T, Robert G, Macfarlane F, Bate P, Kyriakidou O. Diffusion of innovations in service organizations: systematic review and recommendations. Milbank Q 2004;82(4):581-629 [FREE Full text] [doi: 10.1111/j.0887-378X.2004.00325.x] [Medline: 15595944]

35. Aarons GA, Hurlburt M, Horwitz SM. Advancing a conceptual model of evidence-based practice implementation in public service sectors. Adm Policy Ment Health 2011 Jan;38(1):4-23 [FREE Full text] [doi: 10.1007/s10488-010-0327-7] [Medline: 21197565]
36. Glaser B, Strauss A. The constant comparative method of qualitative analysis. Soc Probl 1967;12(4):436-445. [doi: 10.4324/9780203793206-6]
37. Hsieh H, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
38. Saldana J. The Coding Manual For Qualitative Researchers. Thousand Oaks: Sage Publications Ltd; 2015.
39. Guba E. Criteria for assessing the trustworthiness of naturalistic inquiries. Educ Technol Res Dev 1981;29(2):75. [doi: 10.1007/BF02766777]
40. Guba EG, Lincoln Y. Naturalistic Inquiry. Thousand Oaks: SAGE Publications; 1985.
41. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
42. Dedoose. 2015. URL: https://www.dedoose.com/ [accessed 2019-09-25]
43. Kao H, Conant R, Soriano T, McCormick W. The past, present, and future of house calls. Clin Geriatr Med 2009 Feb;25(1):19-34, v. [doi: 10.1016/j.cger.2008.10.005] [Medline: 19217490]
44. Centers for Medicare & Medicaid Services. 2019. CMS Advances Interoperability & Patient Access to Health Data through New Proposals URL: https://www.cms.gov/newsroom/fact-sheets/cms-advances-interoperability-patient-access-health-data-through-new-proposals [accessed 2019-04-06] [WebCite Cache ID 77R8ujduV]
45. Roberts V. Managing strategic outsourcing in the healthcare industry. J Healthc Manag 2001;46(4):239-249. [Medline: 11482242]
46. de Quinn JB. National Chung Hsing University. Outsourcing Innovation: The New Engine of Growth URL: http://web.nchu.edu.tw/pweb/users/arborfish/lesson/8714.pdf [accessed 2019-04-06] [WebCite Cache ID 77R95uzZg]
47. Shinkman R. Outsourcing on the upswing. Health providers are farming out more services to spend less money. Mod Healthc 2000 Sep 4;30(37):46-50, 52. [Medline: 11186524]
48. Quinn J. Strategic outsourcing: leveraging knowledge capabilities. MIT Sloan Manag Rev 1999;40(4):9 [FREE Full text]
49. Green ME. Beware and prepare: The government workforce of the future. Public Pers Manage 2000 Dec;29(4):435-444. [doi: 10.1177/009102600002900402]
50. Anderson MC. Aprimer in measuring outsourcing results. Natl Prod Rev 1997;17(1):33-41. [doi: 10.1002/npr.4040170107]
51. Antonucci Y, Lordi F, Tucker IJ. The pros and cons of IT outsourcing. J Accountancy 1998;185(6):26 [FREE Full text]
52. Bryce DJ, Useem M. The impact of corporate outsourcing on company value. Eur Manag J 1998 Dec;16(6):635-643. [doi: 10.1016/S0263-2373(98)00040-1]
53. Meckbach G. Pros not the whole story: watch out for the cons. Comput Canada 1998;24(11):25-26.
54. Hendry J. Culture, community and networks: The hidden cost of outsourcing. Eur Manage J 1995 Jun;13(2):193-200. [doi: 10.1016/0263-2373(95)00007-8]
55. Welch JA, Nayak PR. Strategic sourcing: a progressive approach to the make-or-buy decision. Acad Manage Perspect 1992 Feb;6(1):23-31. [doi: 10.5465/ame.1992.4274302]
56. Prencipe A. Technological competencies and product's evolutionary dynamics a case study from the aero-engine industry. Res Policy 1997 Jan;25(8):1261-1276. [doi: 10.1016/S0048-7333(96)00900-6]
57. Paoli M, Prencipe A. The role of knowledge bases in complex product systemsome empirical evidence from the aero engine industry. J Manag Gov 1999;3(2):137-160. [doi: 10.1023/A:1009999823012]
58. Willcocks LP, Currie WL. Information technology in public services: towards the contractual organization? Br J Manag 1997 Jun;8(s1):107-120. [doi: 10.1111/1467-8551.8.s1.9]
59. Lonsdale C. Effectively managing vertical supply relationships: a risk management model for outsourcing. Supp Chain Mnagmnt 1999 Oct;4(4):176-183. [doi: 10.1108/13598549910284499]
60. Yu J, Mink PJ, Huckfeldt PJ, Gildemeister S, Abraham JM. Population-level estimates of telemedicine service provision using an all-payer claims database. Health Aff (Millwood) 2018 Dec;37(12):1931-1939. [doi: 10.1377/hlthaff.2018.05116] [Medline: 30633676]
61. Kim J, Park H. Development of a health information technology acceptance model using consumers' health behavior intention. J Med Internet Res 2012 Oct 1;14(5):e133 [FREE Full text] [doi: 10.2196/jmir.2143] [Medline: 23026508]
62. Emani S, Yamin CK, Peters E, Karson AS, Lipsitz SR, Wald JS, et al. Patient perceptions of a personal health record: a test of the diffusion of innovation model. J Med Internet Res 2012 Nov 5;14(6):e150 [FREE Full text] [doi: 10.2196/jmir.2278] [Medline: 23128775]
63. Hsieh JA, Rai A, Keil M. Understanding digital inequality: comparing continued use behavioral models of the socio-economically advantaged and disadvantaged. MIS Quart 2008;32(1):97-126. [doi: 10.2307/25148830]
64. LeRouge C, van Slyke C, Seale D, Wright K. Baby boomers' adoption of consumer health technologies: survey on readiness and barriers. J Med Internet Res 2014 Sep 8;16(9):e200 [FREE Full text] [doi: 10.2196/jmir.3049] [Medline: 25199475]

65. Park J, Erikson C, Han X, Iyer P. Are state telehealth policies associated with the use of telehealth services among underserved populations? Health Aff (Millwood) 2018 Dec;37(12):2060-2068. [doi: 10.1377/hlthaff.2018.05101] [Medline: 30633679]
66. LeRouge C, Garfield MJ. Crossing the telemedicine chasm: have the US barriers to widespread adoption of telemedicine been significantly reduced? Int J Environ Res Public Health 2013 Nov 28;10(12):6472-6484 [FREE Full text] [doi: 10.3390/ijerph10126472] [Medline: 24287864]
67. Flannery D, Jarrin R. Building a regulatory and payment framework flexible enough to withstand technological progress. Health Aff (Millwood) 2018 Dec;37(12):2052-2059. [doi: 10.1377/hlthaff.2018.05151] [Medline: 30633665]
68. LeRouge C, Garfield MJ, Collins RW. Telemedicine: technology mediated service relationship, encounter, or something else? Int J Med Inform 2012 Sep;81(9):622-636. [doi: 10.1016/j.ijmedinf.2012.04.001] [Medline: 22579395]
69. Wingfield N, Thomas K, Abelson R. NJLTC Marketplace. 2018 Jan. Amazon, Berkshire Hathaway and JPMorgan Team Up to Try to Disrupt Health Care URL: http://www.njltc.com/documents/NYTimes%20Amazon%20BH.pdf [accessed 2019-09-24] [WebCite Cache ID 77R9ZGqDb]

Abbreviations

EMR: electronic medical record
VCC: virtual urgent care clinic

Edited by G Eysenbach; submitted 06.04.19; peer-reviewed by L Uscher-Pines, KL Mauco; comments to author 20.06.19; revised version received 11.08.19; accepted 30.08.19; published 15.11.19.

Please cite as:
Sterling R, LeRouge C
On-Demand Telemedicine as a Disruptive Health Technology: Qualitative Study Exploring Emerging Business Models and Strategies Among Early Adopter Organizations in the United States
J Med Internet Res 2019;21(11):e14304
URL: http://www.jmir.org/2019/11/e14304/
doi:10.2196/14304
PMID:31730038

Original Paper

Modeling Research Topics for Artificial Intelligence Applications in Medicine: Latent Dirichlet Allocation Application Study

Bach Xuan Tran[1,2], PhD; Son Nghiem[3], PhD; Oz Sahin[4], PhD; Tuan Manh Vu[5], PhD, MD; Giang Hai Ha[6], MSc; Giang Thu Vu[7], MSc; Hai Quang Pham[6], MD; Hoa Thi Do[8], PhD; Carl A Latkin[2], PhD; Wilson Tam[9], PhD; Cyrus S H Ho[10], MBBS; Roger C M Ho[11,12,13], MBBS

[1]Institute for Preventive Medicine and Public Health, Hanoi Medical University, Hanoi, Vietnam
[2]Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, United States
[3]Centre for Applied Health Economics, Griffith University, Brisbane, Australia
[4]Griffith Climate Change Response Program, Griffith University, Brisbane, Australia
[5]Odonto Stomatology Research Center for Applied Science and Technology, Hanoi Medical University, Hanoi, Vietnam
[6]Institute for Global Health Innovations, Duy Tan University, Da Nang, Vietnam
[7]Center of Excellence in Evidence-based Medicine, Nguyen Tat Thanh University, Ho Chi Minh, Vietnam
[8]Centre of Excellence in Artificial Intelligence in Medicine, Nguyen Tat Thanh University, Ho Chi Minh, Vietnam
[9]Alice Lee Centre for Nursing Studies, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore
[10]Department of Psychological Medicine, National University Hospital, Singapore, Singapore
[11]Center of Excellence in Behavioral Medicine, Nguyen Tat Thanh University, Ho Chi Minh, Vietnam
[12]Department of Psychological Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore
[13]Institute for Health Innovation and Technology, National University of Singapore, Singapore, Singapore

Corresponding Author:
Bach Xuan Tran, PhD
Institute for Preventive Medicine and Public Health
Hanoi Medical University
No 1 Ton That Tung Street
Hanoi, 100000
Vietnam
Phone: 84 98 222 8662
Email: bach.ipmph@gmail.com

Abstract

Background: Artificial intelligence (AI)–based technologies develop rapidly and have myriad applications in medicine and health care. However, there is a lack of comprehensive reporting on the productivity, workflow, topics, and research landscape of AI in this field.

Objective: This study aimed to evaluate the global development of scientific publications and constructed interdisciplinary research topics on the theory and practice of AI in medicine from 1977 to 2018.

Methods: We obtained bibliographic data and abstract contents of publications published between 1977 and 2018 from the Web of Science database. A total of 27,451 eligible articles were analyzed. Research topics were classified by latent Dirichlet allocation, and principal component analysis was used to identify the construct of the research landscape.

Results: The applications of AI have mainly impacted clinical settings (enhanced prognosis and diagnosis, robot-assisted surgery, and rehabilitation), data science and precision medicine (collecting individual data for precision medicine), and policy making (raising ethical and legal issues, especially regarding privacy and confidentiality of data). However, AI applications have not been commonly used in resource-poor settings due to the limit in infrastructure and human resources.

Conclusions: The application of AI in medicine has grown rapidly and focuses on three leading platforms: clinical practices, clinical material, and policies. AI might be one of the methods to narrow down the inequality in health care and medicine between developing and developed countries. Technology transfer and support from developed countries are essential measures for the advancement of AI application in health care in developing countries.

(J Med Internet Res 2019;21(11):e15511) doi:10.2196/15511

KEYWORDS

artificial intelligence; applications; medicine; scientometric; bibliometric; latent Dirichlet allocation

Introduction

The first idea of a thinking machine was developed in 1945 when a system that could amplify human knowledge was described in Vannevar Bush's seminal work [1]. Five years later, Alan Turing mentioned a machine being able to imitate human action and gave chess playing as an example of actions that a computer could do [2]. In 1956, artificial intelligence (AI) was first coined by John McCarthy in a Dartmouth conference [3]. Since then, there have been a few definitions of AI [4-6]. Although there is no consistency in these definitions, one common idea is that AI is an intelligent machine or a system, displaying intelligent behavior.

There are two schools of thought among the AI community: conventional artificial intelligence and computational intelligence [7]. Conventional AI includes machine learning and statistical analysis, while the neural network and fuzzy system belong to computational intelligence [7,8]. Other applications of AI include expert system, automation, and artificial creativity [9]. Expert system and machine learning are two of the most popular applications of AI. The expert system emulates the decision-making ability of humans in a field, while machine learning is a computer program that has the ability to learn from experience. In addition, robotics, a science of dealing with designing and operating robots, with the application of AI, has created robots with improved quality in sensing, vision, and self-awareness [10].

With continuous development and challenges to overcome, AI has been applied in various fields of society such as game playing [11], computer vision [12], speech recognition [13], and expert system in health care [14] and economics [15]. In particular, the contribution of AI in medicine and health care has brought about changes in not only the health system but also patients. The earliest application of AI in medicine dates to 1964, with the corporation of scientists from multidisciplinary research fields for the DENDRAL project [16]. The success of this scientific reasoning is one reason for the explosive spread of AI in biomedicine in the 1970s [17]. Another early application of AI to health care was medical diagnostic decision support systems, which appeared in 1954 [18]. Over the last 60 years, there has been a huge wave of AI technologies in health care. This change is reflected by not only the rapid increase in the number of papers in AI in medicine and health care, but also the appearance of AI in various medical fields [19]. Several AI techniques such as robotics, deep learning, support vector machines, or machine learning have been widely applied in the medical diagnostic system, treatment, and rehabilitation [20-22]. Some scientific publications have shown the effectiveness of AI in medicine and health care. In medical diagnosis, AI has been proved to be effective in improving the diagnostic accuracy for physical diseases [23-25]. The expert system has been used for diagnosis of diseases such as heart disease [26] and diabetes [27] and has proven to be useful for diagnosis and basic treatment advice [27]. For mental illness, AI may be useful for psychiatric consultations. Machine learning has been applied in a predictive model, which could identify patients with symptoms of schizophrenia and attempting to commit suicide with 74% and 80%-90% accuracy, respectively [28,29]. In terms of treatment, most robots assist clinicians in surgery but do not independently perform operations [30].

Due to the variety of AI applications in medicine and health care, there is a need to understand the current states of AI applications, major topics, and the research area of AI in medicine and health care and to identify research gaps. This study attempted to contribute to this understanding by analyzing the context and landscape of research topics [19]. Compared with previous scientometrics research, this study is global and assessed a wide range of AI utilities in medicine and health care [31,32]. Our study used scientific publications downloaded from the Web of Science to model the change and achievement of research topics and landscape in AI applications in health and medicine documents.

Thus, this study evaluated the global development of scientific publications from 1977 to 2018 and characterized research landscapes and constructs of disciplines applied to AI in medicine and health care. By decoding these patterns, we can effectively explore the changes in the growth of publications and may therefore provide better information for other researchers and policymakers in priority settings and evaluation.

Methods

Search Strategies and Data Source

The full strategy of our study has been presented elsewhere [33] (Multimedia Appendix 1). Data were retrieved from the Web of Sciences database provided by Thomson Reuters Institute for Scientific Information. We chose this database because of its outstanding advantages over other databases such as Scopus or PubMed: It contains bibliographic data since 1900, has a higher scientific journal impact, has more indexes, and is better in representing metadata [34].

Data Download

The data under .txt format, including the paper information (publication name, authors, journals' name, year of publication, keywords, author affiliations, total citation, subject research, and abstracts), were downloaded from Web of Science. Two researchers worked independently to simultaneously download the data. Subsequently, we filtered all downloaded data by excluding papers that were published in 2019, not original articles and reviews, written by an anonymous author, and not in English (Multimedia Appendix 2). Any conflict was resolved by discussion. All the data were merged and analyzed by STATA software (STATACorp LLC, College Station, TX).

Data Analysis

We analyzed data based on the characteristic of publication (total papers, publication years, and number of papers by countries), research areas, abstracts (terms and contents of the abstract), citations, and usages (number of downloads). Subsequently, we used STATA software to perform a content analysis of the abstracts. We applied principal component analysis to identify the landscape of AI in medicine and health care. The Jaccard similarity index was utilized to identify research topics or terms most frequently co-occurring with each other. We applied a topic modeling technique for data mining and determining relationships among text documents. Specifically, we chose latent Dirichlet allocation (LDA), which is one of the most popular methods in this field for further analysis. LDA was a helpful technique to classify papers into similar topics [35-39]. It helps recognize the structure of research development, current trends, and interdisciplinary landscapes of research in AI applied to medicine. Using LDA, we classified text in each abstract to a topic where Dirichlet is used as a distribution over discrete distribution; each component in a random vector is the probability of drawing the words/texts associated with that component. Principle component analysis (PCA) was used to classify the research disciplines into corresponding groups.

Thus, by applying LDA, we could obtain an in-depth view of the trends of AI in health care and annotate the documents' topic to discover hidden themes [40]. Additionally, the landscape analysis addressed the relationship between research disciplinaries and showed how research areas in medicine and health care changed due to AI. The summary of analytical techniques for each data types is presented in Table 1.

Table 1. Summary of analytical techniques for each data types.

Type of data and unit of analysis	Analytical methods	Presentations of results
Abstracts		
Words	Frequency of co-occurrence	• Number of papers by countries in abstracts
Papers	Latent Dirichlet allocation	• Ten classifications of research topics
Web of Science classification of research areas		
Web of Science research areas	Coincidence analysis	• Dendrogram of research disciplines (Web of Science classification) • The Web of Science research areas constructing Latent Dirichlet allocation research topics

Results

Number of Published Items and Publication Trend

As seen in Table 1, the number of AI publications increased rapidly during the past 40 years. Notably, most of the publications (23,216 papers, 84.6%) were published during the last 10 years, and 60.6% of the total citation belonged to this period. The usage of papers was counted by the number of downloads. The mean use rate (download rate) within the last 6 months, of papers published in the year 2018 was three times higher than that of papers published in the previous years. The mean use rate within the last 5 years reached its peak in 2013 and decreased from 2012.

We analyzed the frequency of a country where the study was conducted, which was mentioned in the abstract (Table 2). Among 50 countries, the United States appeared the most (1867 times, 40.4%). Notably, only four African countries (Egypt, Niger, Kenya, and Nigeria) were mentioned in the abstracts. In addition, 13 Asian countries contributed to this list, and two Asian leaders of AI technologies—China (including Taiwan and Hong Kong) and India—accounted for 9.8% and 4.32% of the total papers, respectively.

Research Landscapes

Table 3 presents the scientific research topics constructed by LDA. By analyzing the most frequent words and titles, we could manually annotate the label of each topic. Robotics, which most mentioned the 10 topics and branches of AI (topic 1, topic 6, and topic 9), has supported surgery and treatment. AI types were applied the most in the diagnosis and prediction (topic 2, topic 5, and topic 7). Based on development visualization, there was a growing trend in some of the 10 topics, with different rates. The number of papers related to topic 1 was highest and increased gradually but with a slower rate in recent years. Moreover, the number of papers in topic 2 and topic 3 increased at a higher rate than that of papers in other topics (Figure 1).

Table 2. General characteristics of publications.

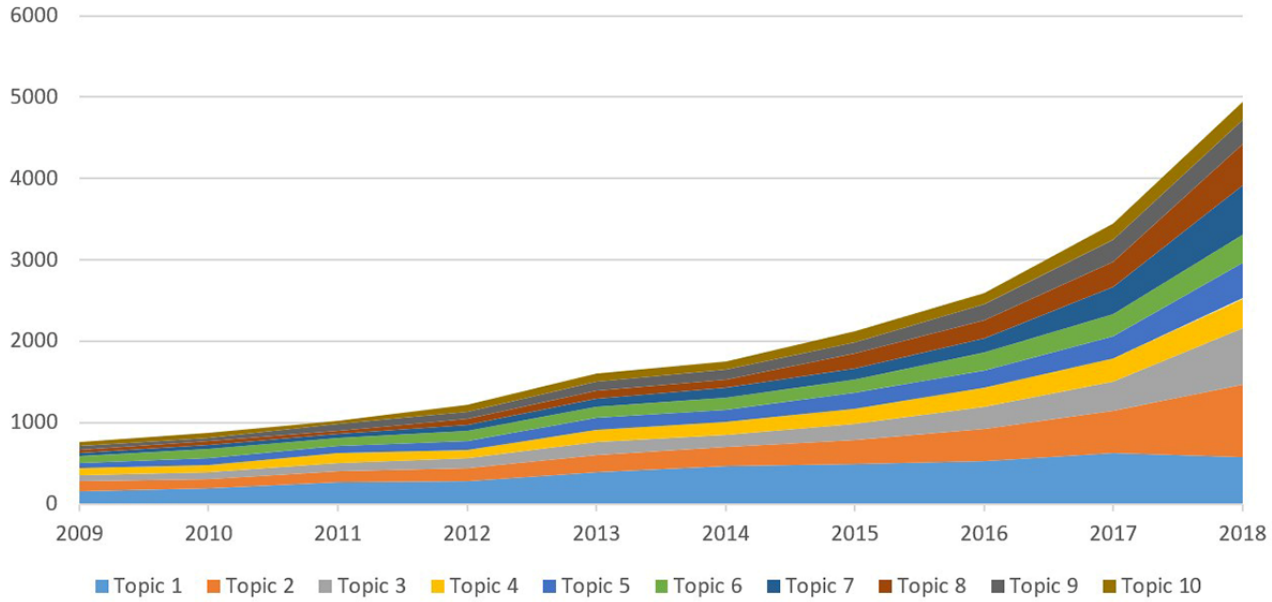
Year published	Total number of papers	Total number of citations	Mean citation rate per year	Total usage in last 6 months	Total usage in last 5 years	Mean use rate in last 6 months	Mean use rate in last 5 years
2018	5619	7084	1.26	35,677	57,370	6.35	2.04
2017	3919	19,639	2.51	11,320	56,890	2.89	2.90
2016	2969	27,636	3.10	6684	55,729	2.25	3.75
2015	2416	31,168	3.23	4,193	46,820	1.74	3.88
2014	1990	30,523	3.07	2473	35,563	1.24	3.57
2013	1839	35,259	3.20	2010	37,934	1.09	4.13
2012	1385	30,130	3.11	1114	19,950	0.80	2.88
2011	1189	37,313	3.92	1379	18,603	1.16	3.13
2010	1010	28,270	3.11	661	10,185	0.65	2.02
2009	880	27,847	3.16	678	9607	0.77	2.18
2008	718	26,865	3.40	530	6944	0.74	1.93
2007	557	19,402	2.90	343	4575	0.62	1.64
2006	479	24,213	3.89	375	4923	0.78	2.06
2005	367	13,460	2.62	178	2473	0.49	1.35
2004	350	16,294	3.10	216	3240	0.62	1.85
2003	262	14,671	3.50	188	2465	0.72	1.88
2002	195	14,143	4.27	157	2109	0.81	2.16
2001	191	8852	2.57	117	1766	0.61	1.85
2000	170	8056	2.49	87	1171	0.51	1.38
1999	150	5517	1.84	61	678	0.41	0.90
1998	163	4396	1.28	44	606	0.27	0.74
1997	124	7179	2.63	89	877	0.72	1.41
1996	114	3310	1.26	29	373	0.25	0.65
1995	98	3182	1.35	38	334	0.39	0.68
1994	100	3570	1.43	37	328	0.37	0.66
1993	61	2238	1.41	29	222	0.48	0.73
1992	62	1395	0.83	25	225	0.40	0.73
1991	41	683	0.59	31	101	0.76	0.49
1990	8	179	0.77	5	16	0.63	0.40
1989	2	438	7.30	2	9	1.00	0.90
1988	7	117	0.54	3	21	0.43	0.60
1987	6	18	0.09	2	8	0.33	0.27
1986	5	59	0.36	4	14	0.80	0.56
1985	2	4	0.06	0	1	0.00	0.10
1984	1	7	0.20	0	3	0.00	0.60
1980	1	51	1.31	0	7	0.00	1.40
1977	1	3	0.07	1	4	1.00	0.80

Table 3. Number of papers by countries as study settings.

Rank	Country settings	Frequency, n (%)
1	United States	1867 (40.4)
2	Ireland	332 (7.2)
3	Taiwan	215 (4.7)
4	China	208 (4.5)
5	United Kingdom	194 (4.2)
6	India	181 (3.9)
7	Japan	164 (3.6)
8	Australia	141 (3.1)
9	Canada	86 (1.9)
10	Iran	83 (1.8)
11	Germany	81 (1.8)
12	Italy	74 (1.6)
13	Brazil	58 (1.3)
14	Spain	56 (1.2)
15	France	55 (1.2)
16	Sweden	43 (0.9)
17	Turkey	43 (0.9)
18	Israel	37 (0.8)
19	New Zealand	33 (0.7)
20	Wallis and Futuna	31 (0.7)
21	Hong Kong	27 (0.6)
22	Mali	27 (0.6)
23	Netherlands	25 (0.5)
24	Poland	25 (0.5)
25	Singapore	23 (0.5)
26	Switzerland	22 (0.5)
27	Greece	21 (0.5)
28	South Africa	20 (0.4)
29	Saudi Arabia	19 (0.4)
30	Malaysia	18 (0.4)
31	Egypt	17 (0.40
32	Pakistan	17 (0.4)
33	Denmark	13 (0.3)
34	Belgium	12 (0.3)
35	Georgia	12 (0.3)
36	Niger	12 (0.3)
37	Kenya	11 (0.2)
38	Mexico	11 (0.2)
39	Nigeria	11 (0.2)
40	Austria	10 (0.2)
41	Finland	10 (0.2)
42	Chile	9 (0.2)

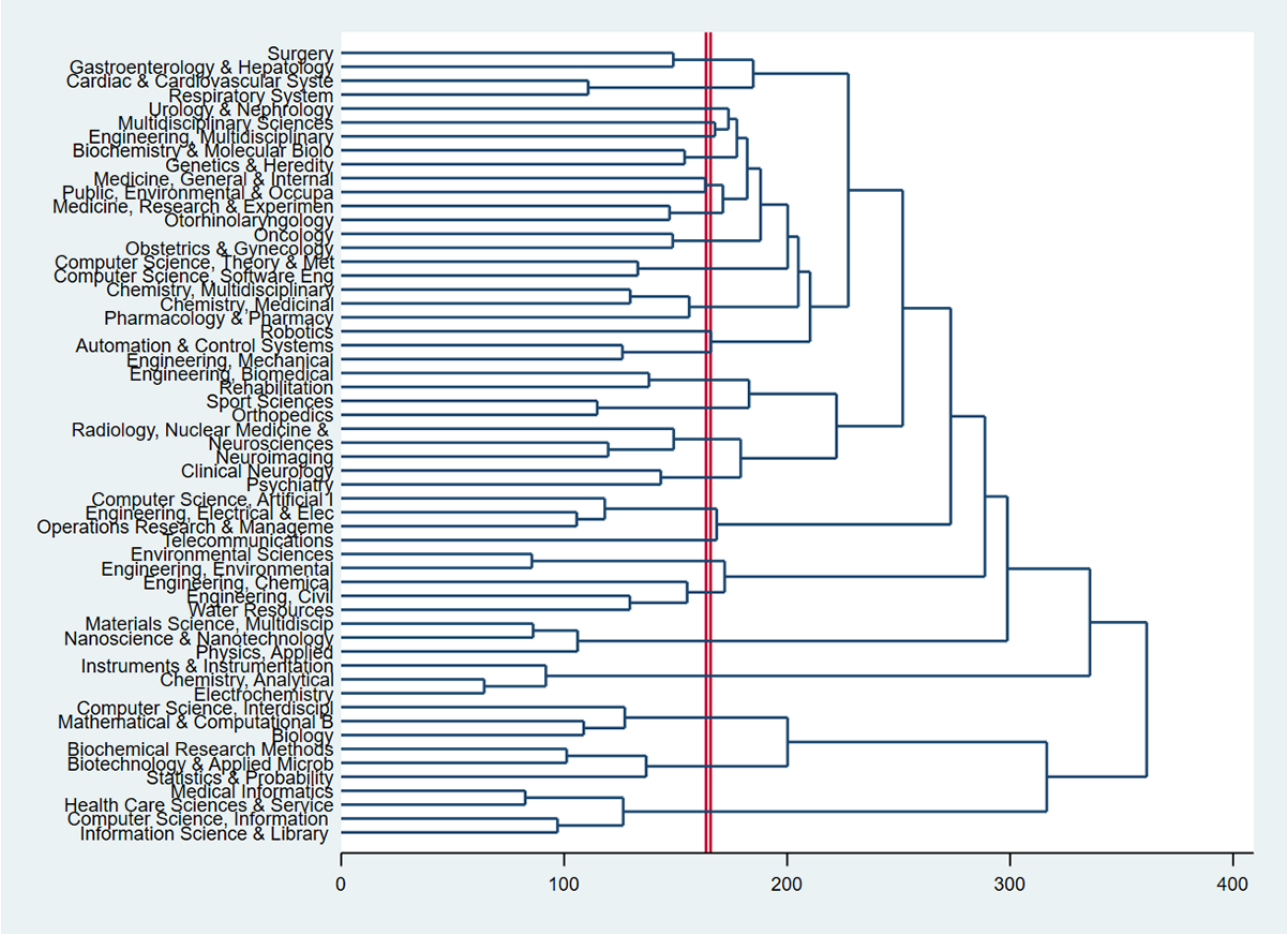
Rank	Country settings	Frequency, n (%)
43	Norway	9 (0.2)
44	Portugal	9 (0.2)
45	Thailand	9 (0.2)
46	United Arab Emirates	9 (0.2)
47	Colombia	8 (0.2)
48	Jordan	8 (0.2)
49	Serbia	8 (0.2)
50	Czech	7 (0.2)

Figure 1. Changes in the applications of artificial intelligence to health and medicine in the past 10 years.



Based on the classification of research areas in the Web of Science, we identified the dendrograms for the areas (Figure 2). The dendrogram includes the clades and leaves. The clade is the branch, and each clade includes one or more research areas. The horizontal axis shows the distance or dissimilarity between research areas. Each joining (fusion) of two clusters is represented on the diagram by the splitting of a vertical line into two vertical lines. The vertical position of the split, shown by a short bar, gives the distance (dissimilarity) between the two research areas. It shows that the AI applications focused on seven following research areas: surgery, robotics, and noncommunicable diseases (hepatocardiac disorders or cancer); neurosciences and psychiatry; the application of electronic health (telecommunication); chemical sciences; nanoscience; electrochemistry; and medical informatics and biotechnology. It seems that AI in medicine was assigned mainly to the disciplines diseases and treatment (surgery or robotics application).

Figure 2. Dendrogram of coincidence of research areas using the Web of Science classifications.



We applied PCA to identify the landscape of AI in medicine and health care (Figure 3). Based on the size of the node, most papers belonged to the following research categories: clinical: surgery, radiology, and nuclear medicine; technology: biomedical, robotics, computer science, medical informatics; and diseases: oncology, general and internal medicine and noncommunicable diseases. As shown in Figure 2, a strong relationship among the applications of AI in treatment, diseases, and medical informatics shows that AI assisted surgeons, especially in some diseases for which surgery is key in treatment or diagnosis, such as cancer or cardiovascular diseases. The combination of information science, computer science, and health care, called health informatics, has created a wide range of applications, from cell level to population level [41]. Collision of several computer science–related fields and medical fields created a multidisciplinary science, which has led to better chances of providing the best treatment to patients. Additionally, the development of computer sciences has contributed to the advancement of AI in pharmacy, biotechnology, and chemistry in areas such as drug discovery, drug identification and validation, and drug trials.

We compared ten research topics by LDA (Table 4) with Web of Science research areas (Multimedia Appendices 3-5) to identify the consistency of research disciplinaries of AI in medicine and health care. Computer science and its related fields appeared the most (eight topics). The major application of computer science has been in medical fields: from cells (gene microbiology information, topic 5), disease (oncology, cardiovascular, topic 7), and diagnosis and treatment (topic 1, topic 6, and topic 9) to health policy (topic 3). Additionally, AI types were used the most in medicine and health care, including expert systems, artificial neural networks, machine learning, and natural language processing. Robot and surgery were two applications mentioned the most in topic 1, topic 6, and topic 9. Robotic-assisted procedures were used for cancer surgery and cardiovascular diseases. Robot-assisted therapy was used in treating sports concussion or neurorehabilitation (topic 9).

Figure 3. Landscapes of artificial intelligence in medicine by Web of Science categories. PCA: principal component analysis.

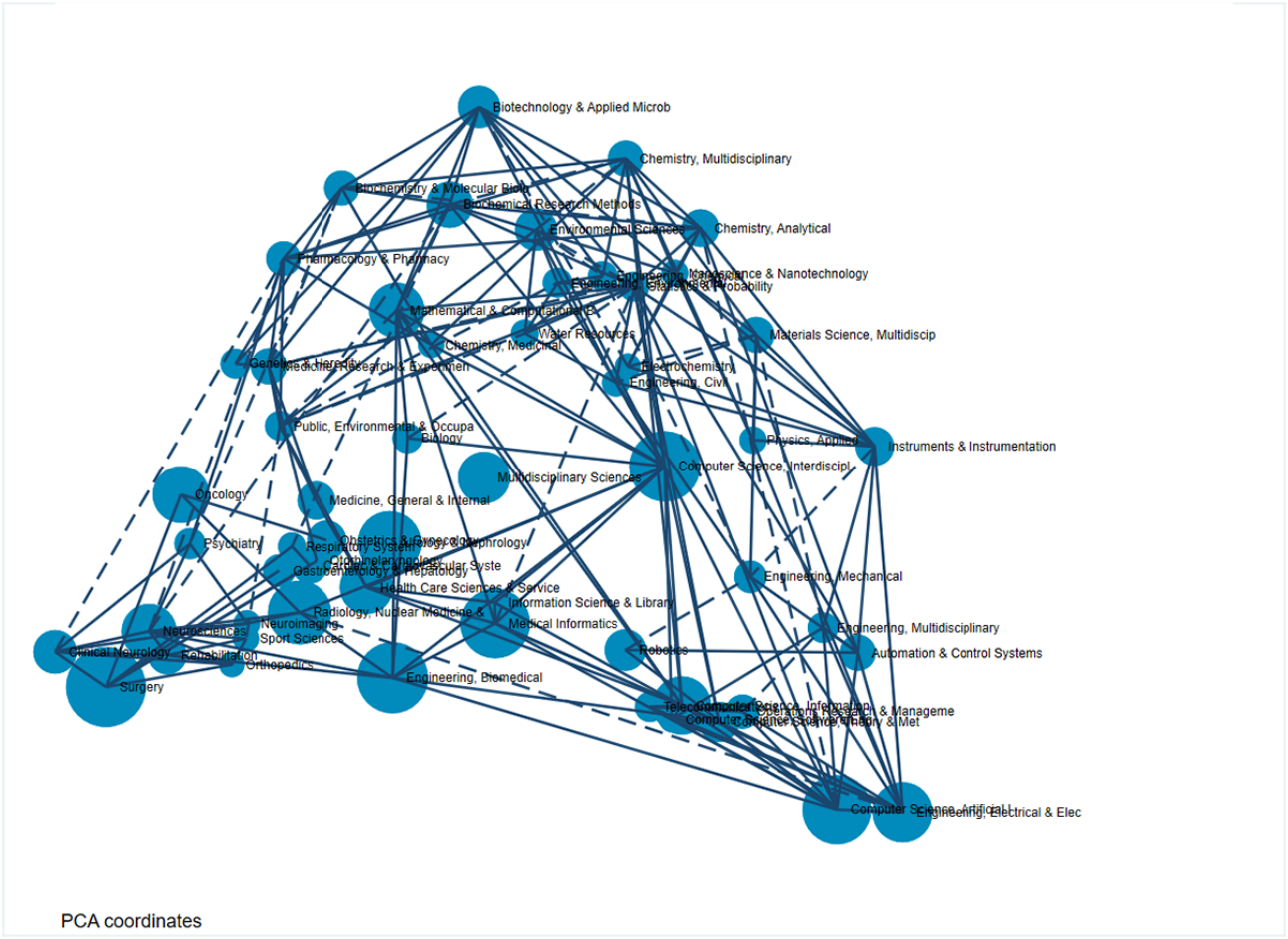


Table 4. Ten research topics classified by latent Dirichlet allocation.

Latent Dirichlet allocation topics	Frequency, n (%)	Topic name
Topic 1	4,3524, 4352 (18.1)	Comparative evaluation of robot-assisted surgery
Topic 2	3662 (15.2)	Expert system for diseases diagnosis and prediction
Topic 3	2839 (11.8)	Health system and policy on AIs[a] in medicine
Topic 4	2182 (9.1)	Artificial neural networks in treatment selection
Topic 5	2089 (8.7)	AI-based gene and protein analysis and prediction
Topic 6	2065 (8.6)	Precision robotics and personalized medicine
Topic 7	2008 (8.4)	Enhanced diagnosis and classification by AI-based images analysis
Topic 8	1922 (8.0)	Using machine learning to predict risk, disease progression, and treatment outcomes
Topic 9	1564 (6.5)	Robot-assisted rehabilitation treatment
Topic 10	1350 (5.6)	Natural language processing tools for biomedical texts and clinical notes

[a]AI: artificial intelligence.

Discussion

Principal Results

By analyzing scientific research publications and modeling their research topics, we generally described the 42-year development and identified the trend of AI application in medicine and health care. The mean use rate related to the application of AI in medicine was the highest in the last 5 years and tended to reduce since 2012. This can be explained by the rapid development of technology and research [42]: Scientific papers published more than 5 years ago would not attract the attention of scientists. Therefore, dissemination efforts need to be taken into consideration by not only policy makers but also authors, to increase the influence and implement changes in practice settings [43]. In addition, the results show a rapid increase in research productivity and downloaded papers in the last 5 years.

Its growth was contributed mostly by western countries, driven by the United States. Among 11 Asian countries in that list, China and India were two leaders in research on AI in medicine. The application of AI has benefitted the health care system in high-income countries. One study showed that the United States could save US $5-$8 billion per year with the application of information technology in health care [44]. Another recent analysis found that with the application of AI, we can save up to US $150 billion in yearly health costs [45]. AI, however, has not been widely used in low-income countries. This could be due to the undeveloped infrastructure in the internet, technology, and health systems and a lack of highly qualified human resource. Regardless of the disadvantages, AI holds promise for changing health care services in low-income countries [46].

Based on the topics and research areas, we found that the application of AI in medicine and health care has been focused on robot support in surgery (topic 1) and rehabilitation (topic 9), AI in diagnosis and clinical decisions support (topic 2, topic 4, topic 5, topic 7, topic 8, and topic 10), and AI in health care system management (topic 3). First, for clinical treatment, our results confirm that medical robots and robot-assisted surgeries have been widely used [47,48]. AI has been widely applied in surgery due to its benefits for patients and medical professionals, such as increased accuracy, reduced operation time, minimized surgical trauma, and reduced length of recovery time for patients [49]. Second, AI methods such as machine learning and natural language process analyze complex medical data [20], decrease time spent finding relevant evidence, and reduce medical errors that improve the quality of diagnosis in medical health care [50]. Finally, AI will certainly be applied more in the health care system in the future owing to its advantages over the traditional decision-making process. On the other hand, the fact that users do not know how the results are analyzed by the "black box" algorithms, ethnic differences in validity of facial recognition technology for genetic diagnosis, medical and behavioral conditions [51,52], and ethnic bias in training data set [53] raise questions about product liability, privacy and data protection, and ethical and legal issues [51]. Thus, researchers have voiced their concern about legacy and ethical guidelines that are lagging behind the development of AI in health care and medicine [51].

Future Implications

Our findings have some implications for health research and policy. The quick development of AI applications in health and medicine requires some preparations. AI may change the relationship between caregivers and patients, as the direct interaction might reduce due to digital tools such as a free app in the patient's personal device, which could diagnose the disease in some cases or even lead to self-diagnosis via the Web [54]. Thus, it is necessary for all parties involved to ensure that, in the case of mental health diagnosis, for instance, subtle signs of mental illness would not be neglected [55]. In addition, standard guidelines or laws about collecting private information or application of AI in all health care sectors are urgently needed [56], as the application of AI in health care and medicine has potential threats to patients' privacy and safety. Finally, AI is transforming health care in resource-poor settings and reducing the gap between rural and urban areas [46]. In rural areas of developing countries, the shortage of medical doctors and trained nurses and the limitation of medical techniques and machines have reduced the quality of medical services [57]. In addition, it is difficult to attract skilled medical workers in rural areas due to the poor working environment and living conditions [58]. However, the development of AI applications can be a solution to these problems. For instance, the AI method (machine learning) proposed a model helping forecast dengue outbreaks in China [59]. In addition, AI has proven to be effective, with a high accuracy of breast cancer detection [60,61]. Moreover, AI can reduce medical costs in developing countries. For example, a highly effective AI method could provide an alternative to expensive diagnostic methods to classify acute leukemia [62]. However, absorptive capacity, local culture, legacy [63], and infrastructure (eg, electricity, internet, or financial source) should be carefully taken into consideration [64]. Notably, policy development for AI should be given more attention, since its failure has been recognized in developing countries such as Vietnam [65].

Limitations

Our study has several limitations worth noting. First, we choose only Web of Science as the database, which may not cover all the publications in the research fields. Second, only English articles and reviews were analyzed in this study. Finally, we applied LDA to model the topic research in title and abstracts, not the full text. However, two other methods (coincidence analysis and PCA) confirmed similar results about the connections of research topics. Thus, LDA could be considered a support method to reduce the workload in the screening step for future systematic reviews [66].

Conclusions

The application of AI in medicine has grown rapidly and focuses on three leading platforms: clinical practices, clinical material, and policies. AI might be one of the methods to reduce the inequality in health care and medicine between developing and developed countries. Technology transfer and support from developed countries, along with the internal efforts of poor-setting countries, help in the development of AI applications in health care and medicine.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search results (Web of Science).

[PDF File (Adobe PDF File), 28 KB - jmir_v21i11e15511_app1.pdf]

Multimedia Appendix 2
Selection of papers in the Web of Science database.
[PDF File (Adobe PDF File), 306 KB - jmir_v21i11e15511_app2.pdf]

Multimedia Appendix 3
The Web of Science research areas constructing latent Dirichlet allocation research topics (topics 1-3).
[PDF File (Adobe PDF File), 25 KB - jmir_v21i11e15511_app3.pdf]

Multimedia Appendix 4
The Web of Science research areas constructing latent Dirichlet allocation research topics (topics 4-6).
[PDF File (Adobe PDF File), 25 KB - jmir_v21i11e15511_app4.pdf]

Multimedia Appendix 5
The Web of Science research areas constructing latent Dirichlet allocation research topics (topics 7-10).
[PDF File (Adobe PDF File), 28 KB - jmir_v21i11e15511_app5.pdf]

References

1. Bush V. As we may think. Interactions 1945;3(2):35-46. [doi: 10.1145/227181.227186]
2. Turing A. Computing machinery and intelligence. Mind 1950;LIX(236):433-460. [doi: 10.1093/mind/LIX.236.433]
3. Nilsson NJ. Stanford University. 2011. Professor John McCarthy URL: https://cs.stanford.edu/memoriam/professor-john-mccarthy [accessed 2019-04-16]
4. High-Level Expert Group on Artificial Intelligence. A definition of Artificial Intelligence: main capabilities and scientific disciplines. Brussels: European Commission; Apr 08, 2019.
5. Ramesh A, Kambhampati C, Monson J, Drew P. Artificial intelligence in medicine. Ann R Coll Surg Engl 2004 Sep 01;86(5):334-338. [doi: 10.1308/147870804290]
6. McCarthy J. Stanford University. 2007. What is AI?/Basic Questions URL: http://jmc.stanford.edu/artificial-intelligence/what-is-ai/index.html [accessed 2019-04-16]
7. Das S, Dey A, Pal A, Roy N. Applications of Artificial Intelligence in Machine Learning: Review and Prospect. IJCA 2015 Apr 22;115(9):31-41. [doi: 10.5120/20182-2402]
8. Bittermann M. Design Computing and Cognition (DCC10). 2010. Artificial Intelligence (AI) versus Computational Intelligence (CI) for treatment of complexity in design Design Computing and Cognition URL: https://pdfs.semanticscholar.org/de17/414444750423573b60bfd206d0047a90c0fa.pdf [accessed 2019-09-06]
9. Singla J. The Diagnosis of Some Lung Diseases in a Prolog Expert System. IJCA 2013 Sep 18;78(15):37-40. [doi: 10.5120/13603-1435]
10. Perez J, Deligianni F, Ravi D, Yang GZ. Artificial intelligence and robotics. London: UK-RAS; 2018.
11. Harbing L. Harvard University. 2017. AI in Video Games: Toward a More Intelligent Game URL: http://sitn.hms.harvard.edu/flash/2017/ai-video-games-toward-intelligent-game/ [accessed 2019-04-16]
12. Gao J, Yang Y, Lin P, Park DS. Computer Vision in Healthcare Applications. J Healthc Eng 2018;2018:5157020 [FREE Full text] [doi: 10.1155/2018/5157020] [Medline: 29686826]
13. Smadi TA, Al Issa HA, Trad E, Smadi KAA. Artificial Intelligence for Speech Recognition Based on Neural Networks. JSIP 2015;06(02):66-72. [doi: 10.4236/jsip.2015.62006]
14. McCauley N, Ala M. The use of expert systems in the healthcare industry. Information & Management 1992 Apr;22(4):227-235 [FREE Full text] [doi: 10.1016/0378-7206(92)90025-b]
15. John McCarthy. 2018. What is AI?/Applications of AI URL: http://jmc.stanford.edu/artificial-intelligence/what-is-ai/applications-of-ai.html [accessed 2019-04-16]
16. Lindsay R, Buchanan B, Feigenbaum E, Lederberg J. Applications of Artificial Intelligence for Organic Chemistry: The DENDRAL Project. New York, NY: McGraw-Hill; 1980.
17. Patel VL, Shortliffe EH, Stefanelli M, Szolovits P, Berthold MR, Bellazzi R, et al. The coming of age of artificial intelligence in medicine. Artif Intell Med 2009 May;46(1):5-17 [FREE Full text] [doi: 10.1016/j.artmed.2008.07.017] [Medline: 18790621]
18. Miller RA. Medical diagnostic decision support systems--past, present, and future: a threaded bibliography and brief commentary. J Am Med Inform Assoc 1994;1(1):8-27 [FREE Full text] [Medline: 7719792]
19. Tran B, Vu G, Ha G, Vuong Q, Ho M, Vuong T, et al. Global Evolution of Research in Artificial Intelligence in Health and Medicine: A Bibliometric Study. JCM 2019 Mar 14;8(3):360. [doi: 10.3390/jcm8030360]
20. Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, et al. Artificial intelligence in healthcare: past, present and future. Stroke Vasc Neurol 2017 Dec;2(4):230-243 [FREE Full text] [doi: 10.1136/svn-2017-000101] [Medline: 29507784]

21. Lundberg SM, Nair B, Vavilala MS, Horibe M, Eisses MJ, Adams T, et al. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. Nat Biomed Eng 2018 Oct;2(10):749-760. [doi: 10.1038/s41551-018-0304-0] [Medline: 31001455]
22. Son Y, Kim H, Kim E, Choi S, Lee S. Application of support vector machine for prediction of medication adherence in heart failure patients. Healthc Inform Res 2010 Dec;16(4):253-259 [FREE Full text] [doi: 10.4258/hir.2010.16.4.253] [Medline: 21818444]
23. Gargeya R, Leng T. Automated Identification of Diabetic Retinopathy Using Deep Learning. Ophthalmology 2017 Dec;124(7):962-969. [doi: 10.1016/j.ophtha.2017.02.008] [Medline: 28359545]
24. Long E, Lin H, Liu Z, Wu X, Wang L, Jiang J, et al. An artificial intelligence platform for the multihospital collaborative management of congenital cataracts. Nat Biomed Eng 2017 Jan 30;1(2). [doi: 10.1038/s41551-016-0024]
25. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 2017 Dec 02;542(7639):115-118. [doi: 10.1038/nature21056] [Medline: 28117445]
26. Al-Shayea Q. Artificial Neural Networks in Medical Diagnosis. International Journal of Computer Science Issues 2011;8(2):150-154 [FREE Full text]
27. Zeki T, Malakooti M, Ataeipoor Y, Tabibi S. An Expert System for Diabetes Diagnosis. American Academic & Scholarly Research Journal 2012;4(5):1-13 [FREE Full text]
28. Gheiratmand M, Rish I, Cecchi GA, Brown MRG, Greiner R, Polosecki PI, et al. Learning stable and predictive network-based patterns of schizophrenia and its clinical symptoms. NPJ Schizophr 2017 May 16;3(1):22. [doi: 10.1038/s41537-017-0022-8] [Medline: 28560268]
29. Just MA, Pan L, Cherkassky VL, McMakin DL, Cha C, Nock MK, et al. Machine learning of neural representations of suicide and emotion concepts identifies suicidal youth. Nat Hum Behav 2017 Oct 30;1(12):911-919. [doi: 10.1038/s41562-017-0234-y]
30. Loh E. Medicine and the rise of the robots: a qualitative review of recent advances of artificial intelligence in health. Leader 2018 Jun 01;2(2):59-63. [doi: 10.1136/leader-2018-000071]
31. Darko A, Chan A, Huo X, Owusu-Manu D. A scientometric analysis and visualization of global green building research. Building and Environment 2019 Feb;149:501-511 [FREE Full text] [doi: 10.1016/j.buildenv.2018.12.059]
32. Ma GP. The Development and Research Trends of Artificial Intelligence in Neuroscience: A Scientometric Analysis in CiteSpace. AMR 2013 Jul;718-720:2068-2073. [doi: 10.4028/www.scientific.net/amr.718-720.2068]
33. Tran BX, Vu GT, Ha GH, Vuong Q, Ho M, Vuong T, et al. Global Evolution of Research in Artificial Intelligence in Health and Medicine: A Bibliometric Study. J Clin Med 2019 Mar 14;8(3) [FREE Full text] [doi: 10.3390/jcm8030360] [Medline: 30875745]
34. Chadegani AA, Salehi H, Yunus MM, Farhadi H, Fooladi M, Farhadi M, et al. A Comparison between Two Main Academic Literature Collections: Web of Science and Scopus Databases. ASS 2013 Apr 27;9(5). [doi: 10.5539/ass.v9n5p18]
35. Li Y, Rapkin B, Atkinson TM, Schofield E, Bochner BH. Leveraging Latent Dirichlet Allocation in processing free-text personal goals among patients undergoing bladder cancer surgery. Qual Life Res 2019 Feb 23:23 [FREE Full text] [doi: 10.1007/s11136-019-02132-w] [Medline: 30798421]
36. Valle D, Albuquerque P, Zhao Q, Barberan A, Fletcher RJ. Extending the Latent Dirichlet Allocation model to presence/absence data: A case study on North American breeding birds and biogeographical shifts expected from climate change. Glob Chang Biol 2018 Dec;24(11):5560-5572 [FREE Full text] [doi: 10.1111/gcb.14412] [Medline: 30058746]
37. Chen C, Zare A, Trinh HN, Omotara GO, Cobb JT, Lagaunne TA. Partial Membership Latent Dirichlet Allocation for Soft Image Segmentation. IEEE Trans Image Process 2017 Dec;26(12):5590-5602. [doi: 10.1109/TIP.2017.2736419] [Medline: 28792897]
38. Lu H, Wei C, Hsiao F. Modeling healthcare data using multiple-channel latent Dirichlet allocation. J Biomed Inform 2016 Apr;60:210-223 [FREE Full text] [doi: 10.1016/j.jbi.2016.02.003] [Medline: 26898516]
39. Gross A, Murthy D. Modeling virtual organizations with Latent Dirichlet Allocation: a case for natural language processing. Neural Netw 2014 Oct;58:38-49. [doi: 10.1016/j.neunet.2014.05.008] [Medline: 24930023]
40. Tong Z, Zhang H. A Document Exploring System on LDA Topic Model for Wikipedia Articles. IJMA 2016 Aug 30;8(3/4):01-13. [doi: 10.5121/ijma.2016.8401]
41. Kukafka R, O'Carroll PW, Gerberding JL, Shortliffe EH, Aliferis C, Lumpkin JR, et al. Issues and opportunities in public health informatics: a panel discussion. J Public Health Manag Pract 2001 Nov;7(6):31-42. [Medline: 11710167]
42. SFR-IA Group, CERF, French Radiology Community. Artificial intelligence and medical imaging 2018: French Radiology Community white paper. Diagn Interv Imaging 2018 Nov;99(11):727-742. [doi: 10.1016/j.diii.2018.10.003] [Medline: 30470627]
43. Brownson R, Eyler A, Harris J, Moore J, Tabak R. Getting the Word Out. Journal of Public Health Management and Practice 2018;24(2):102-111. [doi: 10.1097/phh.0000000000000673]
44. Neumann PJ, Parente ST, Paramore LC. Potential savings from using information technology applications in health care in the United States. Int J Technol Assess Health Care 1996;12(3):425-435. [Medline: 8840663]
45. Collier M, Fu R, Yin L. Accenture. 2017. Artifcial intelligence: healthcare's new nervous system URL: https://www.accenture.com/au-en/insight-artificial-intelligence-healthcare [accessed 2019-04-16]

46. Wahl B, Cossy-Gantner A, Germann S, Schwalbe NR. Artificial intelligence (AI) and global health: how can AI contribute to health in resource-poor settings? BMJ Glob Health 2018 Aug 29;3(4):e000798. [doi: 10.1136/bmjgh-2018-000798] [Medline: 30233828]
47. Leung T, Vyas D. Robotic Surgery: Applications. Am J Robot Surg 2014 Jun 01;1(1):1-64 [FREE Full text] [Medline: 26501128]
48. Nishimura K. Current status of robotic surgery in Japan. Korean J Urol 2015 Mar;56(3):170-178 [FREE Full text] [doi: 10.4111/kju.2015.56.3.170] [Medline: 25763120]
49. Kose E, Ozturk NN, Karahan SR. Artificial Intelligence in Surgery. Eur Arch Med Res 2018 Dec 26;34(Suppl 1):4-6. [doi: 10.5152/eamr.2018.43043]
50. Walczak S. The Role of Artificial Intelligence in Clinical Decision Support Systems and a Classification Framework. International Journal of Computers in Clinical Practice 2018:31-47. [doi: 10.4018/IJCCP.2018070103]
51. Rigby M. Ethical Dimensions of Using Artificial Intelligence in Health Care. AMA Journal of Ethics 2019;21(2):121-124. [doi: 10.1001/amajethics.2019.121]
52. Martinez-Martin N. What Are Important Ethical Implications of Using Facial Recognition Technology in Health Care? AMA J Ethics 2019 Mar 01;21(2):E180-E187 [FREE Full text] [doi: 10.1001/amajethics.2019.180] [Medline: 30794128]
53. Dolgin E. AI face-scanning app spots signs of rare genetic disorders. Nature 2019 Jan 7:A. [doi: 10.1038/d41586-019-00027-x]
54. White RW, Horvitz E. Web to world: predicting transitions from self-diagnosis to the pursuit of local medical assistance in web search. AMIA Annu Symp Proc 2010;2010:882-886 [FREE Full text] [Medline: 21347105]
55. Artificial Intelligence in Healthcare. London: Academy of Medical Royal Colleges; 2019.
56. Mesko B. The role of artificial intelligence in precision medicine. Expert Review of Precision Medicine and Drug Development 2017 Sep 20;2(5):239-241. [doi: 10.1080/23808993.2017.1380516]
57. Guo J, Li B. The Application of Medical Artificial Intelligence Technology in Rural Areas of Developing Countries. Health Equity 2018;2(1):174-181 [FREE Full text] [doi: 10.1089/heq.2018.0037] [Medline: 30283865]
58. Mullei K, Mudhune S, Wafula J, Masamo E, English M, Goodman C, et al. Attracting and retaining health workers in rural areas: investigating nurses' views on rural posts and policy interventions. BMC Health Serv Res 2010 Jul 02;10 Suppl 1:S1 [FREE Full text] [doi: 10.1186/1472-6963-10-S1-S1] [Medline: 20594367]
59. Guo P, Liu T, Zhang Q, Wang L, Xiao J, Zhang Q, et al. Developing a dengue forecast model using machine learning: A case study in China. PLoS Negl Trop Dis 2017 Oct;11(10):e0005973 [FREE Full text] [doi: 10.1371/journal.pntd.0005973] [Medline: 29036169]
60. Rodriguez-Ruiz A, Lång K, Gubern-Merida A, Broeders M, Gennaro G, Clauser P, et al. Stand-Alone Artificial Intelligence for Breast Cancer Detection in Mammography: Comparison With 101 Radiologists. J Natl Cancer Inst 2019 Mar 05:5. [doi: 10.1093/jnci/djy222] [Medline: 30834436]
61. Houssami N, Kirkpatrick-Jones G, Noguchi N, Lee CI. Artificial Intelligence (AI) for the early detection of breast cancer: a scoping review to assess AI's potential in breast screening practice. Expert Rev Med Devices 2019 May;16(5):351-362. [doi: 10.1080/17434440.2019.1610387] [Medline: 30999781]
62. Escalante HJ, Montes-y-Gómez M, González JA, Gómez-Gil P, Altamirano L, Reyes CA, et al. Acute leukemia classification by ensemble particle swarm model selection. Artif Intell Med 2012 Jul;55(3):163-175. [doi: 10.1016/j.artmed.2012.03.005] [Medline: 22510477]
63. Sher P, Wong V, Shaw V, Hooley G, Loveridge R, Wilson D. Absorptive Capacity and Learning in Technology Transfer: The Case of Taiwanese Information Technology Firms. Internationalization 1998:105-121. [doi: 10.1007/978-1-349-26556-5_5]
64. Adepoju IOO, Albersen BJA, De Brouwere V, van Roosmalen J, Zweekhorst M. mHealth for Clinical Decision-Making in Sub-Saharan Africa: A Scoping Review. JMIR Mhealth Uhealth 2017 Mar 23;5(3):e38 [FREE Full text] [doi: 10.2196/mhealth.7185] [Medline: 28336504]
65. Vuong QH. The (ir)rational consideration of the cost of science in transition economies. Nat Hum Behav 2018 Jan;2(1):5. [doi: 10.1038/s41562-017-0281-4] [Medline: 30980055]
66. Mo Y, Kontonatsios G, Ananiadou S. Supporting systematic reviews using LDA-based document representations. Syst Rev 2015 Nov 26;4(1):172. [doi: 10.1186/s13643-015-0117-0] [Medline: 26612232]

Abbreviations

AI: artificial intelligence
LDA: latent Dirichlet allocation
PCA: principal component analysis

Edited by G Eysenbach; submitted 16.07.19; peer-reviewed by HP Do, H Vuong; comments to author 14.08.19; revised version received 28.08.19; accepted 28.08.19; published 01.11.19.

Please cite as:
Tran BX, Nghiem S, Sahin O, Vu TM, Ha GH, Vu GT, Pham HQ, Do HT, Latkin CA, Tam W, Ho CSH, Ho RCM
Modeling Research Topics for Artificial Intelligence Applications in Medicine: Latent Dirichlet Allocation Application Study
J Med Internet Res 2019;21(11):e15511
URL: https://www.jmir.org/2019/11/e15511
doi:10.2196/15511
PMID:31682577

Original Paper

Performance of Fetal Medicine Foundation Software for Pre-Eclampsia Prediction Upon Marker Customization: Cross-Sectional Study

Karina Bilda De Castro Rezende[1,2,3*], MD, MSc; Antonio José Ledo Alves Cunha[1,4*], MD, PhD; Joffre Amim Jr[2,3*], MD, PhD; Wescule De Moraes Oliveira[5*], MD; Maria Eduarda Belloti Leão[2*], MD; Mariana Oliveira Alves Menezes[2*], MD; Ana Alice Marques Ferraz De Andrade Jardim[2,3*], MD; Rita Guérios Bornia[2,3*], MD, PhD

[1]Programa de Pós Graduação em Clínica Médica, Faculdade de Medicina, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil

[2]Maternidade Escola, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil

[3]Programa de Mestrado Profissional em Saúde Perinatal, Maternidade Escola, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil

[4]Laboratório Multidisciplinar de Epidemiologia e Saúde -LAMPES, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil

[5]Faculdade de Medicina, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil

[*]all authors contributed equally

Corresponding Author:
Karina Bilda De Castro Rezende, MD, MSc
Maternidade Escola
Universidade Federal do Rio de Janeiro
Rua das Laranjeiras,180
Laranjeiras
Rio de Janeiro, 22240-000
Brazil
Phone: 55 2122857935
Email: karina@me.ufrj.br

Abstract

Background: FMF2012 is an algorithm developed by the Fetal Medicine Foundation (FMF) to predict pre-eclampsia on the basis of maternal characteristics combined with biophysical and biochemical markers. Afro-Caribbean ethnicity is the second risk factor, in magnitude, found in populations tested by FMF, which was not confirmed in a Brazilian setting.

Objective: This study aimed to analyze the performance of pre-eclampsia prediction software by customization of maternal ethnicity.

Methods: This was a cross-sectional observational study, with secondary evaluation of data from FMF first trimester screening tests of singleton pregnancies. Risk scores were calculated from maternal characteristics and biophysical markers, and they were presented as the risk for early pre-eclampsia (PE34) and preterm pre-eclampsia (PE37). The following steps were followed: (1) identification of women characterized as black ethnicity; (2) calculation of early and preterm pre-eclampsia risk, reclassifying them as white, which generated a new score; (3) comparison of the proportions of women categorized as high risk between the original and new scores; (4) construction of the receiver operator characteristic curve; (5) calculation of the area under the curve, sensitivity, and false positive rate; and (6) comparison of the area under the curve, sensitivity, and false positive rate of the original with the new risk by chi-square test.

Results: A total of 1531 cases were included in the final sample, with 219 out of 1531 cases (14.30; 95% CI 12.5-16.0) and 182 out of 1531 cases (11.88%; 95% CI 10.3-13.5) classified as high risk for pre-eclampsia development, originally and after recalculating the new risk, respectively. The comparison of FMF2012 predictive model performance between the originally estimated risks and the estimated new risks showed that the difference was not significant for sensitivity and area under the curve, but it was significant for false positive rate.

Conclusions: We conclude that black ethnicity classification of Brazilian pregnant women by the FMF2012 algorithm increases the false positive rate. Suppressing ethnicity effect did not improve the test sensitivity. By modifying demographic characteristics, it is possible to improve some performance aspects of clinical prediction tests.

(J Med Internet Res 2019;21(11):e14738) doi:10.2196/14738

KEYWORDS

decision support techniques; mass screening; pre-eclampsia; ethnicity; algorithms

Introduction

Pre-Eclampsia

Pre-eclampsia is predominant in gestational hypertensive disorders, with a significant impact on maternal and neonatal health [1,2]. Many researchers aim to identify pre-eclampsia development in high-risk pregnancies by using an effective predictive model. This would allow the implementation of strategies for efficient prevention of disease occurrence in a selected population, thereby reducing its prevalence [3]. At present, there is no customized model for clinical use in pregnant Brazilian women.

The Fetal Medicine Foundation Software

FMF2012 [4] is an algorithm, developed by the Fetal Medicine Foundation (FMF) to predict pre-eclampsia on the basis of maternal characteristics combined with biophysical and biochemical markers is available on the FMF website at no cost [5]. It estimates the likelihood of developing pre-eclampsia from maternal factors (ethnicity/skin color, age, weight, height, history of diabetes, chronic hypertension, autoimmune diseases, and use of assisted reproduction techniques), together with biophysical markers such as mean arterial pressure and uterine artery pulsatility index (UtAPI) [5,6]. The objective of screening in the first trimester is to identify women at high risk for preterm pre-eclampsia (<37 weeks) and reduce such a risk through prophylactic use of low-dose aspirin.

Brazilian Experience

In 2016, this FMF2012 model was tested in a sample of pregnant Brazilian women, and its performance was found to be unsatisfactory because of differences in the contribution of risk factors such as ethnicity/skin color [7,8]. According to the FMF, the screening-positive rate in black women is greater than that in white women, as an inevitable consequence of the fact that the prevalence of preterm pre-eclampsia is more than three time higher in black than in white women [9].

Afro-Caribbean ethnicity is the second most common risk factor identified in populations tested by the FMF [10], which was not confirmed in our population. The variable, maternal ethnicity, applied in the FMF2012 algorithm overestimates the risk, which compromises the performance of the screening, as it is a variable with a coefficient of great magnitude [7].

We proposed to suppress the effect of ethnicity on the risk estimated for pre-eclampsia in the sample studied.

Objectives

The objective of this study was to analyze the performance of pre-eclampsia prediction software by customization of maternal ethnicity in a Brazilian scenario.

Methods

Study Design

This was a cross-sectional observational study, with secondary evaluation of data from first trimester screening tests of single-fetus pregnancies performed between October 2010 and December 2015.

Setting

The study was conducted at the Maternidade Escola da Universidade Federal do Rio de Janeiro, a nonprofit university hospital that exclusively serves patients from the public health system and receives undergraduate and postgraduate students in the health care sector. The local ethics committee approved the study protocol (CAAE: 78764117.0.0000.5275).

The following exclusion criteria were the same as described by the FMF and applied to the original study [8]: pregnancy with chromosomal or structural abnormality, miscarriage or fetal death before 24 weeks of gestation, use of acetylsalicylic acid (ASA) during pregnancy before 16 weeks of gestation, and delivery of a small-for-gestational-age newborn to a mother without pre-eclampsia.

First Trimester Screening Scan

Patients were scheduled for a first trimester screening scan at 11+0 to 13+6 weeks of gestation. This examination included recording of maternal characteristics, measurement of fetal crown-rump length, measurement of right and left UtAPIs by transabdominal color Doppler ultrasound (Nemio, Toshiba; Xario, Toshiba; Medison V10, Medison; or Aloka, Aloka Co), and measurement of mean arterial pressure with an automated device (3BTO-A2, Microlife or OMRON, OMRON Corporation) by using a standardized method (in both arms simultaneously, while the mother was sitting after ≥10-min rest) [11]. All data were entered into the FMF2012 software.

In the FMF software, the lists in risk calculation for ethnicity classification have been fixed, and the mother should be categorized using the following [5]: (1) white (European, Middle Eastern, North African, and Hispanic), (2) black (African, Caribbean, and African American), (3) East Asian (Chinese, Japanese, and Korean), (4) South Asian (Indian, Pakistani, and Bangladeshi), (5) mixed (white-black, white–East Asian, white–South Asian, black–East Asian, black–South Asian, and East Asian–South Asian). However, in Brazil and in our study, the criterion to define ethnicity was self-qualification of skin color [1,12].

The maternal characteristics were collected from a patient questionnaire administered by a medical doctor. Continuous variables were maternal age (in years), weight (in kilograms), and height (in centimeters). Categorical variables were self-reported ethnicity (black, white, yellow, indigenous, or mixed) [12], parity (nulliparous, parous with no previous

pre-eclampsia, or parous with previous pre-eclampsia), maternal family history of pre-eclampsia (yes or no), smoking during pregnancy (yes or no), history of previous hypertension (yes or no), type 1 diabetes (yes or no), type 2 diabetes (yes or no), systemic lupus erythematosus or antiphospholipid syndrome (yes or no), and use of assisted reproductive technology (yes or no).

The biophysical markers considered in this study were crown-rump length (in millimeters), mean arterial pressure (in mm Hg and multiples of median) [11], and mean UtAPIs (arithmetic mean and in multiples of median) [13]. The FMF2012 algorithm calculated multiples of median values by using a multiples of median equation [6].

Risk scores were calculated according to the competitive risk model described by Wright et al [6] from maternal characteristics and biomarkers (mean arterial pressure and UtAPIs), and these were presented as the risk of pre-eclampsia development before 34 and 37 weeks. The cut-off values for positivity for these timepoints were 1/200 and 1/57, respectively [14].

Data on pregnancy outcomes (pre-eclampsia occurrence and gestational age at delivery) were collected from hospital records. The diagnosis of early pre-eclampsia was based on the onset of systolic blood pressure≥140 mm Hg or diastolic blood pressure≥90 mm Hg and proteinuria (protein excretion>300 mg/24 hours after 20 weeks of gestation), which requires delivery before 34 weeks (pre-eclampsia<34 weeks or early pre-eclampsia) or before 37 weeks (pre-eclampsia<37 weeks or preterm pre-eclampsia) [14,15]. Gestational age at birth was calculated on the basis of the last menstrual period or first trimester ultrasound screening. When the difference between these timepoints was >7 days, ultrasound estimation was used.

For our purpose, the following process was undertaken: (1) identification of pregnant women characterized as ethnically black from recalled first trimester reports; (2) calculation of risk of early pre-eclampsia and preterm pre-eclampsia by the FMF2012 algorithm in these pregnant women, reclassifying them as white, which generates a new score; (3) comparison of the proportion of women categorized as high risk by the original and new scores; (4) construction of the receiver operator characteristic curve; (5) calculation of the area under the curve (AUC), sensitivity, and false-positive rate (FPR) and respective 95% CIs; and (6) comparison of the AUC, sensitivity, and FPR of the original risk with the *new* risks by using a Chi-square test (the differences were considered significant if $P<.05$). STATA 13 statistical software package (StataCorp, College Station, Texas) was used for data analyses.

Results

First trimester screening was carried out in 1934 singleton pregnancies. We excluded 403 cases because of fetal aneuploidies (n=7); major fetal malformation (n=28); miscarriage, termination, or fetal death before 24 weeks of gestation (n=18); ASA use at ≤16 weeks of gestation (n=103); small-for-gestational-age neonatal status in the absence of pre-eclampsia (n=69); and missing outcome data (n=178). The remaining 1531 cases were included in the study. We identified 645 (645/1531, 42.12%) patients classified as mixed, 589 (589/1531, 38.47%) as white, and 296 (296/1531, 19.33%) as ethnically black. The sample presented 11 (0.71%) cases of early pre-eclampsia and 26 (1.69%) cases of preterm pre-eclampsia. We observed that 3 of 11 cases (27%) of early pre-eclampsia and 6 of 26 (23%) cases of preterm pre-eclampsia occurred in pregnant women primarily classified as ethnically black.

Mean maternal weight, height, and age were 67 kg, 160 cm, and 27 years, respectively. According to the predetermined cut-off values, 219 of 1531 cases (14.30%, 95% CI 12.5-16.0) of our final sample were classified to be at high risk of pre-eclampsia development. After we recalculated the *new risk*, 182 of 1531 cases (11.88%, 95% CI 10.3-13.5) of the final sample were categorized as being at high risk of pre-eclampsia development.

The pre-eclampsia rate in this sample was not different in relation to ethnicity, smoking, family history of pre-eclampsia, or assisted reproductive technology use; as our sample contained few cases of the latter, significant inference was not drawn in this case. In addition, no case of systemic lupus erythematosus or antiphospholipid syndrome was detected in our sample.

Table 1 presents an evaluation of the performance of the FMF2012 predictive model among the studied population, according to the originally estimated risks, with pregnant black women classified as ethnically black, and the *newly estimated risks*, which consider all patients as white/mixed race (baseline risk) for pre-eclampsia<34 weeks and pre-eclampsia<37 weeks. The comparison of the FMF2012 predictive model performance between the originally estimated risks and the *newly estimated risks* showed that the difference was not significant for sensitivity, but it was significant for FPR.

Figures 1 and 2 present the AUC and the comparison between receiver operator characteristic curves of the original risk and *new risk* for pre-eclampsia<34 weeks and pre-eclampsia<37 weeks, respectively. There were no significant differences between the curves.

Table 1. Results of the evaluation of the performance of the Fetal Medicine Foundation 2012 (FMF2012) predictive model.

Outcome	Sensitivity (%), (95% CI)	*P* value	False-positive (%), (95% CI)	*P* value	Area under curve (95% CI)	*P* value
Pre-eclampsia <34 weeks						
Original risk	63 (35-92)	.66	13.9 (12.2-15.6)	.05[a]	0.84 (0.71-0.97)	.17
New risk	54 (25-89)	.66	11.5 (10-13)	.05[a]	0.80 (0.65-0.96)	.17
Pre-eclampsia <37 weeks						
Original risk	46 (27-65)	.57	13.9 (12.2-15.7)	.04[a]	0.77 (0.68-0.86)	.36
New risk	38 (20-57)	.57	11.5 (10-13)	.04[a]	0.76 (0.65-0.85)	.36

[a]*P*≤.05.

Figure 1. Receiver operator characteristic curves of the original risk and new risk for pre-eclampsia <34. ROC: receiver operator characteristic.

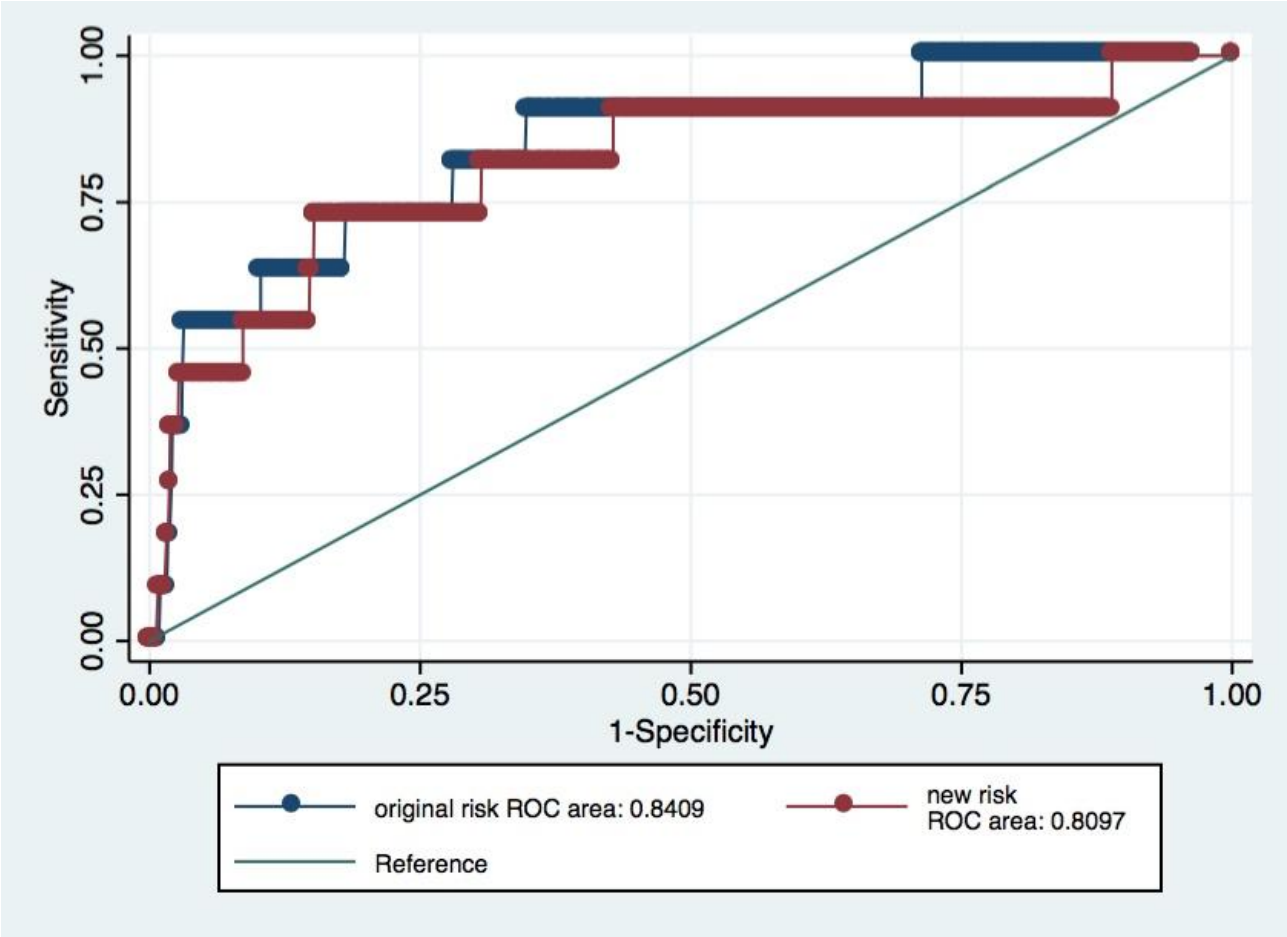
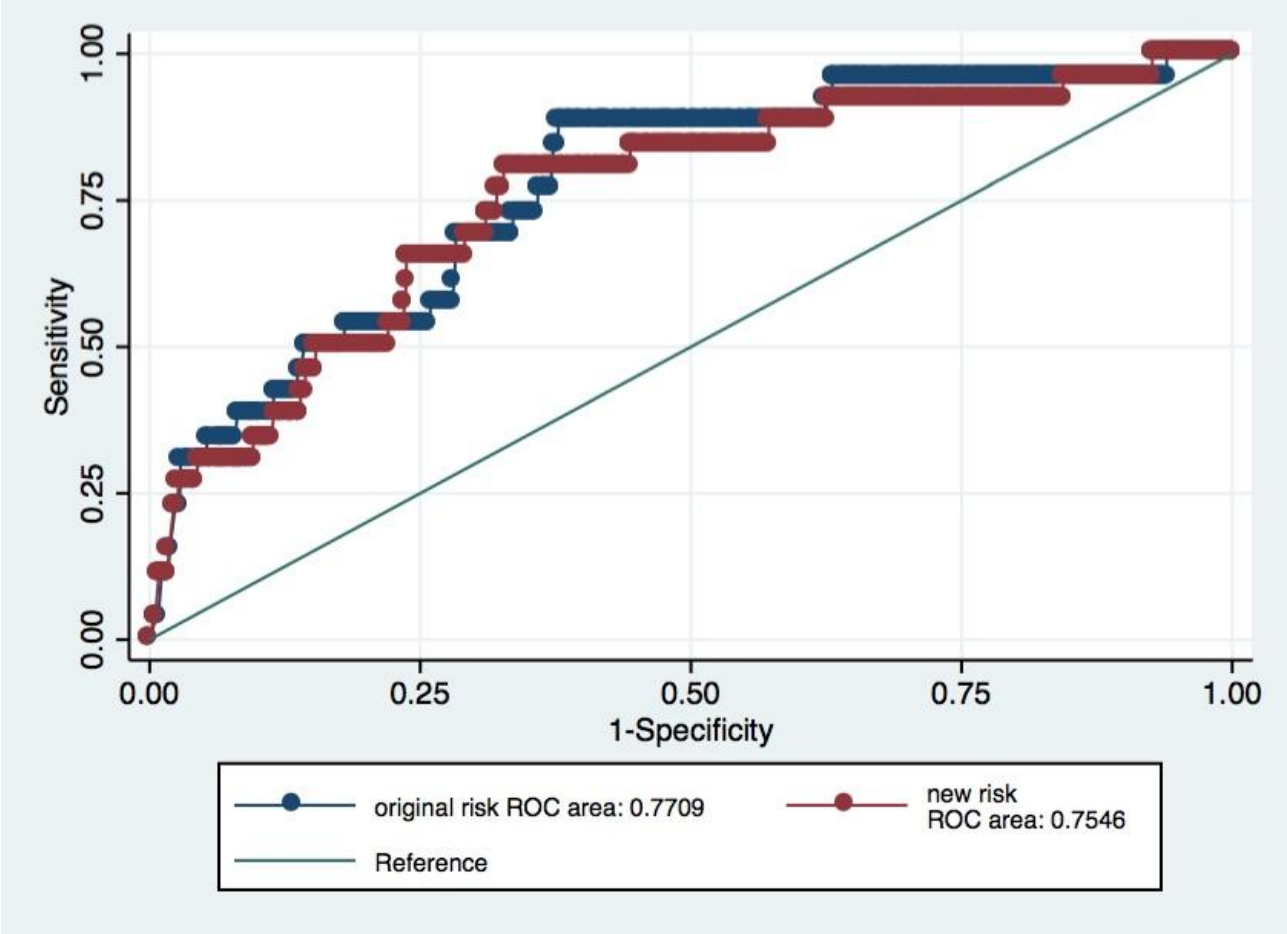


Figure 2. Receiver operator characteristic curves of the original risk and new risk for pre-eclampsia <37. ROC: receiver operator characteristic.



Discussion

Principal Findings

This study proposed a new strategy to use the predictive model of pre-eclampsia, FMF2012, in the first trimester of gestation in Brazilian women, to optimize its application; a customized model for our population is not available as yet. However, pre-eclampsia prediction was not improved by the suggested strategy, as the sensitivity remained the same. A simulation that all pregnant women who were submitted to pre-eclampsia screening in the first trimester, along the study period, would belong to the white race was applied, which represents the baseline risk of the predictive model, suppressing the effect of race on the model. It was found that a significantly lower proportion of screened pregnant women would be categorized as high risk with this approach, which implies a reduction in the FPR.

According to the FMF2012 algorithm for pre-eclampsia risk assessment, maternal racial origin is a categorical variable with the following possible values: *white*, *black*, *East Asian*, *South Asian*, or *mixed*; only black and South Asian ethnicity demonstrate significant contribution for the prediction of pre-eclampsia [6,16,17].

Ethnic disparities remain to be a contentious matter [18]. The racial classification applied in our country by Instituto Brasileiro de Geografia e Estatística identifies people with regard to their race, and it is also used in national administrative databases [19]. Biological methods based on the identification of biogeographical ancestry are not suited for the intended purposes, and the racial composition obtained by self-classification seems to be the most accurate because of historical and theoretical reasons [20].

In this study, the criterion to define ethnicity was self-qualification of skin color, which constitutes one of the characteristics that comprises ethnicity that is not associated with ancestry, which, in turn, contributes to the validation of this attribute in the model, as it is not the racial determinant, especially in the mixed Brazilian population [20]. The reclassification of black patients as ethnically white creates a new score that is less than the original score, as being ethnically white denotes the baseline risk and eliminates the differences regarding the risk of pre-eclampsia development. FMF published that the prevalence of preterm pre-eclampsia is more than three times higher in black than in white women, which was not observed in our sample, which results in a greater FPR.

Even with a small number of early pre-eclampsia cases in the studied sample, a significant decrease in the FPR was quantified. However, there was no significant difference in the sensitivity and AUC, which dictates the performance of diagnostic tests. This was also observed in the performance of preterm pre-eclampsia screening, but the number of cases was twice as that in this study. Although, in population terms, the improvement in the FPR was small and the study did not have

a considerable impact on overall detection rates, making allowances for ethnic origin can make a significant difference to an individual patient-specific risk, which could alter clinical decision making, mainly with regard to ASA prescription [21,22]. Furthermore, incorrect classifications of a pregnant woman as high risk could cause her to follow centralized prenatal care with frequent exams and rigorous protocols that could be stressful and adversely impact the financial and social costs.

The limitations of this study are related to the small number of cases of early and preterm pre-eclampsia in the sample and mainly in ethnically black women, and this study addressed only one maternal characteristic included as a predictor factor of pre-eclampsia. This study focused on a regional question regarding the performance of the FMF2012 algorithm in pre-eclampsia screening. Although Brazil is a country of continental dimensions with widespread and social inequalities, there are some islands of quality health assistance that can create effective screening strategies and disseminate them by using training programs.

Conclusions

This study makes an important contribution to the understanding of the effect of black ethnicity in our sample. It also tests an alternative approach that can improve prenatal follow-up and health indicators. It is a work-around approach to employ and take advantage of available software of clinical prediction models. Users are allowed to customize demographic characteristics to adjust predefined coefficients, in different ways, without changing the algorithm structure. This approach can be extended to other characteristics in other algorithms, but the knowledge of the effect of the subject's characteristic as a risk factor on the target population is a key pillar to achieve performance improvement in clinical prediction models with the proposed strategy in different scenarios. In our sample, making allowance for ethnic origin can make a significant difference to an individual patient-specific risk, which could alter clinical decision making. In conclusion, the classification of pregnant Brazilian women as ethnically black by an FMF2012 pre-eclampsia screening test increases the FPR. Suppressing the effect of ethnicity did not improve test sensitivity. By modifying demographic characteristics, it is possible to improve some performance aspects of clinical prediction tests.

Conflicts of Interest

None declared.

References

1. Lenfant C, National Education Program Working Group on High Blood Pressure in Pregnancy. Working group report on high blood pressure in pregnancy. J Clin Hypertens (Greenwich) 2001;3(2):75-88 [FREE Full text] [doi: 10.1016/0002-9378(90)90653-O] [Medline: 11416689]
2. World Health Organization. 2011. Who Recommendations for Prevention and Treatment of Pre-Eclampsia and Eclampsia. URL: https://extranet.who.int/rhl/guidelines/who-recommendations-prevention-and-treatment-pre-eclampsia-and-eclampsia [accessed 2019-05-16]
3. Rolnik DL, Wright D, Poon LC, O'Gorman N, Syngelaki A, Matallana C, et al. Aspirin versus placebo in pregnancies at high risk for preterm preeclampsia. N Engl J Med 2017 Aug 17;377(7):613-622. [doi: 10.1056/NEJMoa1704559] [Medline: 28657417]
4. Dumitras AM. Scribd. 2012 Dec. The First Trimester Screening Module 2012 Algorithm. URL: https://pt.scribd.com/document/137707519/03-Guide-for-the-First-Trimester-Screening-M [accessed 2018-12-01]
5. The Fetal Medicine Foundation. URL: https://courses.fetalmedicine.com/user/get_license/ [accessed 2019-05-16]
6. Wright D, Akolekar R, Syngelaki A, Poon LC, Nicolaides KH. A competing risks model in early screening for preeclampsia. Fetal Diagn Ther 2012;32(3):171-178. [doi: 10.1159/000338470] [Medline: 22846473]
7. Rezende KB, Cunha AJ, Pritsivelis C, Faleiro EC, Junior J, Bornia RG. How do maternal factors impact preeclampsia prediction in Brazilian population? J Matern Fetal Neonatal Med 2017 Nov;32(7):1051-1056. [doi: 10.1080/14767058.2017.1399115] [Medline: 29082782]
8. Rezende KB, Cunha AJ, Junior J, Bornia RG. External validation of the Fetal Medicine Foundation algorithm for the prediction of preeclampsia in a Brazilian population. Pregnancy Hypertens 2019 Jul;17:64-68. [doi: 10.1016/j.preghy.2019.05.006] [Medline: 31487659]
9. The Fetal Medicine Foundation. URL: https://fetalmedicine.org/ [accessed 2019-08-01]
10. Khalil A, Rezende J, Akolekar R, Syngelaki A, Nicolaides KH. Maternal racial origin and adverse pregnancy outcome: a cohort study. Ultrasound Obstet Gynecol 2013 Mar;41(3):278-285 [FREE Full text] [doi: 10.1002/uog.12313] [Medline: 23023978]
11. Tayyar A, Krithinakis K, Wright A, Wright D, Nicolaides KH. Mean arterial pressure at 12, 22, 32 and 36 weeks' gestation in screening for pre-eclampsia. Ultrasound Obstet Gynecol 2016 May;47(5):573-579 [FREE Full text] [doi: 10.1002/uog.15815] [Medline: 26582336]
12. Osório R. Educational Action. 2003. O Sistema Classificatório De 'Cor Ou Raça' Do Ibge. URL: http://www.acaoeducativa.org.br/fdh/wp-content/uploads/2012/10/quesito-cor-IBGE.pdf [accessed 2018-09-21]

13. Plasencia W, Maiz N, Bonino S, Kaihura C, Nicolaides KH. Uterine artery Doppler at 11 + 0 to 13 + 6 weeks in the prediction of pre-eclampsia. Ultrasound Obstet Gynecol 2007 Oct;30(5):742-749 [FREE Full text] [doi: 10.1002/uog.5157] [Medline: 17899573]
14. Akolekar R, Syngelaki A, Poon L, Wright D, Nicolaides KH. Competing risks model in early screening for preeclampsia by biophysical and biochemical markers. Fetal Diagn Ther 2013;33(1):8-15 [FREE Full text] [doi: 10.1159/000341264] [Medline: 22906914]
15. Brown MA, Lindheimer MD, de Swiet M, van Assche A, Moutquin JM. The classification and diagnosis of the hypertensive disorders of pregnancy: statement from the International Society for the Study of Hypertension in Pregnancy (ISSHP). Hypertens Pregnancy 2001;20(1):IX-XIV. [doi: 10.1081/PRG-100104165] [Medline: 12044323]
16. Wright D, Syngelaki A, Akolekar R, Poon LC, Nicolaides KH. Competing risks model in screening for preeclampsia by maternal characteristics and medical history. Am J Obstet Gynecol 2015 Jul;213(1):62.e1-62e10. [doi: 10.1016/j.ajog.2015.02.018] [Medline: 25724400]
17. Tan MY, Syngelaki A, Poon LC, Rolnik DL, O'Gorman N, Delgado JL, et al. Screening for pre-eclampsia by maternal factors and biomarkers at 11-13 weeks' gestation. Ultrasound Obstet Gynecol 2018 Aug;52(2):186-195 [FREE Full text] [doi: 10.1002/uog.19112] [Medline: 29896812]
18. Crawford NG, Kelly DE, Hansen ME, Beltrame MH, Fan S, Bowman SL, NISC Comparative Sequencing Program, et al. Loci associated with skin pigmentation identified in African populations. Science 2017 Nov 17;358(6365):pii: eaan8433 [FREE Full text] [doi: 10.1126/science.aan8433] [Medline: 29025994]
19. The Brazilian Institute of Geography and Statistics. 1997. Minimum Social Indicators - ISM. URL: https://www.ibge.gov.br/estatisticas/sociais/educacao/17374-indicadores-sociais-minimos.html?=&t=notas-tecnicas [accessed 2017-05-26]
20. Santos DJ, Palomares NB, Normando D, Quintão CC. Raça versus etnia: diferenciar para melhor aplicar. Dental Press J Orthod 2010 Jun;15(3):121-124. [doi: 10.1590/s2176-94512010000300015]
21. Spencer K, Heath V, El-Sheikhah A, Ong CY, Nicolaides KH. Ethnicity and the need for correction of biochemical and ultrasound markers of chromosomal anomalies in the first trimester: a study of Oriental, Asian and Afro-Caribbean populations. Prenat Diagn 2005 May;25(5):365-369. [doi: 10.1002/pd.1153] [Medline: 15906426]
22. Watt HC, Wald NJ, Smith D, Kennard A, Densem J. Effect of allowing for ethnic group in prenatal screening for Down's syndrome. Prenat Diagn 1996 Aug;16(8):691-698. [doi: 10.1002/(SICI)1097-0223(199608)16:8<691::AID-PD946>3.0.CO;2-2] [Medline: 8878277]

Abbreviations

ASA: acetylsalicylic acid
AUC: area under the curve
FMF: Fetal Medicine Foundation
FPR: false-positive rate
UtAPIs: uterine artery pulsatility index

Edited by G Eysenbach; submitted 20.05.19; peer-reviewed by C Sauders, J Rezende Filho; comments to author 18.07.19; revised version received 02.09.19; accepted 02.09.19; published 22.11.19.

Please cite as:
Rezende KBDC, Cunha AJLA, Amim Jr J, Oliveira WDM, Leão MEB, Menezes MOA, Jardim AAMFDA, Bornia RG
Performance of Fetal Medicine Foundation Software for Pre-Eclampsia Prediction Upon Marker Customization: Cross-Sectional Study
J Med Internet Res 2019;21(11):e14738
URL: http://www.jmir.org/2019/11/e14738/
doi:10.2196/14738
PMID:31755874

Commentary

The Connection Between the Nervous System and Machines: Commentary

Giacomo Valle[1,2], MSc

[1]The Biorobotics Institute, Sant'Anna School of Advanced Studies, Pisa, Italy

[2]Translational Neural Engineering Lab, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

Corresponding Author:
Giacomo Valle, MSc
The Biorobotics Institute
Sant'Anna School of Advanced Studies
Viale Rinaldo Piaggio 34
Pisa,
Italy
Phone: 39 3405454807
Email: vallegiacomo@gmail.com

Related Article:

Comment on: http://jmir.org/2019/10/e16194/

Abstract

Decades of technological developments have populated the field of brain-machine interfaces and neuroprosthetics with several replacement strategies, neural modulation treatments, and rehabilitation techniques to improve the quality of life for patients affected by sensory and motor disabilities. This field is now quickly expanding thanks to advances in neural interfaces, machine learning techniques, and robotics. Despite many clinical successes, and multiple innovations in animal models, brain-machine interfaces remain mainly confined to sophisticated laboratory environments indicating a necessary step forward in the used technology. Interestingly, Elon Musk and Neuralink have recently presented a new brain-machine interface platform with thousands of channels, fast implantation, and advanced signal processing. Here, how their work takes part in the context of the restoration of sensory-motor functions through neuroprostheses is commented.

(J Med Internet Res 2019;21(11):e16344) doi:10.2196/16344

KEYWORDS

brain-machine interfaces; neural electrodes; neural recording; neurostimulation; sensory-motor dysfunctions

Significant research in biology, medicine and engineering has sought to obtain effective solutions to improve quality of life of human subjects affected by sensory-motor disorders. Neuroprosthetics are implantable devices designed to replace or improve the function of a disabled part of the nervous system [1]. This technology is relatively recent, as the first neuroprosthetic device successfully implanted was a cochlear implant in 1957 [2]. Since then, such an approach has been expanded to many different applications, which include motor prosthetics [3-6], sensorimotor prosthetics [7-9], visual prosthetics [10,11], and cognitive prosthetics [12].

Up till now, patients who used brain-machine interfaces have had a quite poor perception of the instantaneous behavior, position, or motion of the robotic device, which has prevented them from operating in fully closed-loop and natural control. The restoration of sensory feedback and voluntary control, along with the development and successful integration of these sensor modalities, is a mandatory step towards the realization of future bidirectional neuroprostheses [13].

The challenges described above can be addressed by creating a brain-machine interface that utilizes the processing power of the human brain to control the robotic device. Directly connecting to the human nervous system means closing the gap between user intent and the expected behavior of the apparatus. Furthermore, generating a shorter loop between user intent and device behavior or motion (by eliminating part of the low-level sensor-based control) will allow for easier control, a reduced learning investment, and a reduced cognitive burden of operating the device.

Neural interfaces play a pivotal role in the efficacy of a neuroprosthetic. Due to their ability to read out electrical activity from the nervous system, it is possible to decode signals into cognitive, sensory, or motor information through the use of computational methods. This information can then be used to control a prosthetic device, robot, or computer. It also induces better understanding of brain behavior through the recording of neural activity, providing information about sensory areas responsible for hearing or sight (sensory prosthetics), or helping to regulate malfunctioning motor functions (motor prosthetics). On the other hand, pacemaker or bladder control neuroprosthetics also use similar physical principles, targeting the autonomic nervous system and helping patients with paraplegia due to spinal cord damage [14].

In a recent article, Elon Musk and his company Neuralink presented a new platform to target the brain for neuroprosthetic applications [15]. They used arrays of small and flexible electrodes (called threads), with 3072 electrodes per array, distributed across 96 threads. They also developed a neurosurgical robot able to insert six threads (192 electrodes) per minute. Each thread can be individually inserted into the brain with high precision, avoiding surface vasculature and targeting specific brain zones. The electrode array is packaged into a small implantable device that contains custom chips for low power, onboard amplification, and digitization. Moreover, since neural spikes in a brain-machine interface must be detected in real time to maximize decoding efficacy, Neuralink has developed a custom online spike-detection software that has achieved a spiking yield of up to 70% in chronically implanted electrodes. Musk's long-term idea consists of enabling humans to connect their brains to machines, and Neuralink's approach to a brain-machine interface has shown unprecedented packaging density, extensibility, and scalability in a clinically relevant package. The main properties of the neural electrodes are related to their biocompatibility, long-term stability, and recording/stimulating selectivity when interfacing with both peripheral and central nervous systems [16,17]. Therefore, more tests should be performed for a complete validation of this new platform. This step is not trivial, as it is crucial to show the possible translation of this approach to humans. Further, it is necessary to demonstrate the effective benefits of using this new technology in comparison to other techniques that have been widely tested in the previous decades. The hypothetical complete brain-machine connection has become a closer possibility, but it is not ready just yet.

In this field, many devices and smart materials have been presented as effective solutions to interfacing with nervous tissues, enabling an intimate connection between the brain and machines in animals and even in humans [18]. Understanding how to interact with the brain using advanced algorithms has become of great clinical interest now, both to decode neural information [19] and to encode natural sensations by exploiting biomimetic neurostimulation strategies [20]. Moreover, advanced data processing methods have to be developed to bring these technologies to real life application. In this direction, new tools like machine learning and quantum computing will help to bring this concept to reality.

In the near future, neurotechnologies will continue to grow. More accurate and advanced computer simulations (eg, computational modelling) will allow researchers to test and validate these technologies even quicker. Implantable neurotechnologies will literally become part of us. Direct bidirectional communication between the brain and external devices, the transformation that this connection brings about, and the blurring of the boundaries between humans and machines, are issues that raise several ethical, social, and cultural concerns. Personal identity, physical integrity, and the human dignity [21] of people using the next generation of brain-machine interfaces will surely require further attention.

Conflicts of Interest

None declared.

References

1. Borton D, Micera S, Millán JDR, Courtine G. Personalized neuroprosthetics. Sci Transl Med 2013 Dec 06;5(210):210rv2-210rv2. [doi: 10.1126/scitranslmed.3005968] [Medline: 24197737]
2. Eisen MD. Djourno, Eyries, and the first implanted electrical neural stimulator to restore hearing. Otol Neurotol 2003 May;24(3):500-506. [doi: 10.1097/00129492-200305000-00025] [Medline: 12806307]
3. van den Brand R, Heutschi J, Barraud Q, DiGiovanna J, Bartholdi K, Huerlimann M, et al. Restoring voluntary control of locomotion after paralyzing spinal cord injury. Science 2012 Jul 01;336(6085):1182-1185 [FREE Full text] [doi: 10.1126/science.1217416] [Medline: 22654062]
4. Collinger JL, Wodlinger B, Downey JE, Wang W, Tyler-Kabara EC, Weber DJ, et al. High-performance neuroprosthetic control by an individual with tetraplegia. The Lancet 2013 Feb;381(9866):557-564. [doi: 10.1016/s0140-6736(12)61816-9]
5. Wagner FB, Mignardot J, Le Goff-Mignardot CG, Demesmaeker R, Komi S, Capogrosso M, et al. Targeted neurotechnology restores walking in humans with spinal cord injury. Nature 2018 Nov 31;563(7729):65-71. [doi: 10.1038/s41586-018-0649-2] [Medline: 30382197]
6. Hochberg LR, Bacher D, Jarosiewicz B, Masse NY, Simeral JD, Vogel J, et al. Reach and grasp by people with tetraplegia using a neurally controlled robotic arm. Nature 2012 May 16;485(7398):372-375 [FREE Full text] [doi: 10.1038/nature11076] [Medline: 22596161]

7. Flesher SN, Collinger JL, Foldes ST, Weiss JM, Downey JE, Tyler-Kabara EC, et al. Intracortical microstimulation of human somatosensory cortex. Sci Transl Med 2016 Oct 19;8(361):361ra141-361ra141. [doi: 10.1126/scitranslmed.aaf8083] [Medline: 27738096]
8. Raspopovic S, Capogrosso M, Petrini FM, Bonizzato M, Rigosa J, Di Pino G, et al. Restoring natural sensory feedback in real-time bidirectional hand prostheses. Sci Transl Med 2014 Mar 05;6(222):222ra19-222ra19 [FREE Full text] [doi: 10.1126/scitranslmed.3006820] [Medline: 24500407]
9. Petrini FM, Bumbasirevic M, Valle G, Ilic V, Mijović P, Cvancara P, et al. Sensory feedback restoration in leg amputees improves walking speed, metabolic cost and phantom pain. Nat Med 2019 Sep 9;25(9):1356-1363. [doi: 10.1038/s41591-019-0567-3] [Medline: 31501600]
10. Dobelle WH, Mladejovsky MG, Girvin JP. Artifical vision for the blind: electrical stimulation of visual cortex offers hope for a functional prosthesis. Science 1974 Mar 01;183(4123):440-444. [doi: 10.1126/science.183.4123.440] [Medline: 4808973]
11. da Cruz L, Dorn JD, Humayun MS, Dagnelie G, Handa J, Barale P, Argus II Study Group. Five-Year Safety and Performance Results from the Argus II Retinal Prosthesis System Clinical Trial. Ophthalmology 2016 Oct;123(10):2248-2254 [FREE Full text] [doi: 10.1016/j.ophtha.2016.06.049] [Medline: 27453256]
12. Andersen RA, Hwang EJ, Mulliken GH. Cognitive neural prosthetics. Annu Rev Psychol 2010;61:169-90, C1 [FREE Full text] [doi: 10.1146/annurev.psych.093008.100503] [Medline: 19575625]
13. SensArs - Neuroprosthetics. 2019. URL: https://www.sensars.com/ [accessed 2019-10-26]
14. Creasey GH, Grill JH, Korsten M, Betz R, Anderson R, Walter J, Implanted Neuroprosthesis Research Group. An implantable neuroprosthesis for restoring bladder and bowel control to patients with spinal cord injuries: a multicenter trial. Arch Phys Med Rehabil 2001 Dec;82(11):1512-1519. [doi: 10.1053/apmr.2001.25911] [Medline: 11689969]
15. Musk E, Neuralink. An Integrated Brain-Machine Interface Platform With Thousands of Channels. J Med Internet Res 2019 Oct 31;21(10):e16194 [FREE Full text] [doi: 10.2196/16194] [Medline: 31642810]
16. Navarro X, Krueger TB, Lago N, Micera S, Stieglitz T, Dario P. A critical review of interfaces with the peripheral nervous system for the control of neuroprostheses and hybrid bionic systems. J Peripher Nerv Syst 2005 Oct;10(3):229-258. [doi: 10.1111/j.1085-9489.2005.10303.x] [Medline: 16221284]
17. Lebedev MA, Nicolelis MAL. Brain-machine interfaces: past, present and future. Trends Neurosci 2006 Oct;29(9):536-546. [doi: 10.1016/j.tins.2006.07.004] [Medline: 16859758]
18. Schiavone G, Lacour SP. Conformable bioelectronic interfaces: Mapping the road ahead. Sci Transl Med 2019 Jul 31;11(503):eaaw5858. [doi: 10.1126/scitranslmed.aaw5858] [Medline: 31366582]
19. Schwemmer MA, Skomrock ND, Sederberg PB, Ting JE, Sharma G, Bockbrader MA, et al. Meeting brain-computer interface user performance expectations using a deep neural network decoding framework. Nat Med 2018 Nov;24(11):1669-1676. [doi: 10.1038/s41591-018-0171-y] [Medline: 30250141]
20. Valle G, Mazzoni A, Iberite F, D'Anna E, Strauss I, Granata G, et al. Biomimetic Intraneural Sensory Feedback Enhances Sensation Naturalness, Tactile Sensitivity, and Manual Dexterity in a Bidirectional Prosthesis. Neuron 2018 Oct 10;100(1):37-45.e7 [FREE Full text] [doi: 10.1016/j.neuron.2018.08.033] [Medline: 30244887]
21. Drew L. The ethics of brain-computer interfaces. Nature 2019 Jul 24;571(7766):S19-S21. [doi: 10.1038/d41586-019-02214-2] [Medline: 31341310]

Edited by G Eysenbach; submitted 20.09.19; peer-reviewed by A Korchi; comments to author 12.10.19; revised version received 26.10.19; accepted 31.10.19; published 06.11.19.

Please cite as:
Valle G
The Connection Between the Nervous System and Machines: Commentary
J Med Internet Res 2019;21(11):e16344
URL: https://www.jmir.org/2019/11/e16344
doi:10.2196/16344
PMID:31692449

Viewpoint

Intelligent Sensing to Inform and Learn (InSTIL): A Scalable and Governance-Aware Platform for Universal, Smartphone-Based Digital Phenotyping for Research and Clinical Applications

Scott Barnett[1*], BEng, BSc, PhD; Kit Huckvale[2*], BSc, MBChB, MSc, PhD; Helen Christensen[2,3], BA, MPsychol, PhD; Svetha Venkatesh[1], BTech, MTech, PhD; Kon Mouzakis[2], BSc, MSc; Rajesh Vasa[1], PhD

[1]Applied Artificial Intelligence Institute (A2I2), Deakin University, Geelong, Australia

[2]Black Dog Institute, UNSW Sydney, Randwick, Australia

[3]Mindgardens Neuroscience Network, Sydney, Australia

[*]these authors contributed equally

Corresponding Author:
Kit Huckvale, BSc, MBChB, MSc, PhD
Black Dog Institute
UNSW Sydney
Hospital Road
Randwick, 2031
Australia
Phone: 61 2 9382 4530
Fax: 61 2 9382 8510
Email: c.huckvale@unsw.edu.au

Abstract

In this viewpoint we describe the architecture of, and design rationale for, a new software platform designed to support the conduct of digital phenotyping research studies. These studies seek to collect passive and active sensor signals from participants' smartphones for the purposes of modelling and predicting health outcomes, with a specific focus on mental health. We also highlight features of the current research landscape that recommend the coordinated development of such platforms, including the significant technical and resource costs of development, and we identify specific considerations relevant to the design of platforms for digital phenotyping. In addition, we describe trade-offs relating to data quality and completeness versus the experience for patients and public users who consent to their devices being used to collect data. We summarize distinctive features of the resulting platform, InSTIL (Intelligent Sensing to Inform and Learn), which includes universal (ie, cross-platform) support for both iOS and Android devices and privacy-preserving mechanisms which, by default, collect only anonymized participant data. We conclude with a discussion of recommendations for future work arising from learning during the development of the platform. The development of the InSTIL platform is a key step towards our research vision of a population-scale, international, digital phenotyping bank. With suitable adoption, the platform will aggregate signals from large numbers of participants and large numbers of research studies to support modelling and machine learning analyses focused on the prediction of mental illness onset and disease trajectories.

(J Med Internet Res 2019;21(11):e16399) doi:10.2196/16399

KEYWORDS

eHealth; e-Mental health; mHealth; digital phenotyping; personal sensing; smartphone; iPhone; software development; software framework; technology platform

Introduction

There has been a recent explosion of interest in consumer digital tools across health care, such as smartphones. Mental health has been no exception [1]. Clinically valuable applications have been identified in depression and anxiety, suicidality, drug and alcohol disorders, ageing and dementia, and neurological disease. Through digital phenotyping [2,3] (or personal sensing [4]), behavioral signals, sensor data, and self-reported information gathered through smartphones, wearable sensors and smart home devices can be combined to elucidate the nature and clinical status of health conditions, such as depression [5,6],

anxiety [7], and bipolar disorder [8-13]. These signals also promise better insight into the earliest signs of mental disorders, such as changes in sleep, social behavior, and cognitive function, and raise the prospect of robust, individual-level risk stratification and prediction [14]. Modern smartphones contain multiple sensors that may be used for both passive (ie, continuously running in the background without intervention) and active (ie, with direct input from the user) data collections. Individual sensors may be used for both, such as applying a device accelerometer to collect measurements from a stereotyped movement task while also recording circadian movement patterns.

Importantly, these digital tools can be designed to use gathered data to deliver and enhance interventions [2]. Techniques such as artificial intelligence can transform data into self-management recommendations concerning relapse or treatment adjustment, and can also drive interactive experiences, such as chatbot-based conversations [15,16]. The use of online and smartphone apps for the prevention and management of mood disorders [17], suicide [18], bipolar disorder [19], and the promotion of mental wellbeing has either already been established or is under active investigation and can deliver clinical outcomes comparable to face-to-face therapy [20-24]. Automation has created opportunities to reach global users in a timely way; creating new mechanisms to support health care users that do not rely on face-to-face services [25]. Data-driven tailoring also has the potential to address key challenges of poor engagement and meaningful use that are commonly seen in today's apps [26,27].

Today, efforts to realize a vision of smart sensing and adaptive intervention design are fragmented and often narrowly focused. Despite enthusiastic efforts to build sensing and intervention apps, many appear to have limited potential for clinical translation because they are not backed by evidence of efficacy [28], or do not clearly satisfy clinical quality [29] and safety expectations (such as those concerning diagnostic accuracy [30]). Our own review suggests that the choice of data collected appears more often determined by technical ease rather than what is most important to understand physiological, neurological, and psychological processes [31]. There are also numerous questions of data privacy [32], acceptability [33], and the ethics of data collection [34] which have been fueled by high-profile public data scandals. Left unaddressed, these questions risk undermining public and professional confidence in this transformative technology area. Crucially, without large-scale public participation, much of the vision for digital phenotyping and adaptive interventions cannot be realized. It is becoming increasingly clear that a complete and comprehensive data platform is required to capture the breadth of available sensor data in a meaningful way. This itself is a challenge, given the need to specify the various functional modules required to support the range of available sensors without negative impacts on user experience and to meaningfully curate the data that is produced.

As a field, and at the beginning of what may prove to be one of the most useful ways we can help understand the nature, trajectory, treatment, and mechanisms of mental health disorders, we need to change the way we operate. Over the next few years, if appropriately and ethically acquired, we believe that the time has come to create an international digital phenotype bank. Its purpose will be to explore the relationship between behavior and the development, and subsequent trajectory, of common mental illnesses. It will also create the opportunity to develop personalized digital interventions that assist individuals in identifying and preempting periods of ill health. Because predictive analytics and related methods can efficiently leverage existing data [35], data consolidated in a digital phenotype bank promises a multiplier effect from reuse by clinical, research and, with appropriate governance, industry users.

However, to enable these research and clinical uses, signals data from potentially large numbers of participants must be collected, marshalled, and persisted. A key enabler of this vision is, therefore, the availability of scalable software platforms backed by appropriate technical architecture. This paper describes the design of a new platform for digital phenotyping intended to satisfy this requirement. We outline the design goals of the platform and highlight aspects that researchers may wish to consider when developing similar platforms.

A Case for Shared Platforms

Overview

The technology landscape surrounding the development of both digital phenotype data collection and data-driven intervention studies can be characterized by significant heterogeneity (of technologies, platforms, clinical problems, and research approaches [31]). Nevertheless, several common features stand out. These motivate our interest in the potential for shared data platforms and are summarized below.

Most Digital Data Collection and Intervention Tools Are Custom-Built

Researchers in electronic mental health (e-mental health) often build custom digital solutions to support their studies. This typically involves creating study-specific mobile (or web) apps, often from scratch. These digital health research apps contain code for gathering study specific data from sensors, surveys, and custom workflows (such as games). Additionally, these apps must handle aspects such as user management, data transport to a secure server, data privacy, and ensuring that the technical solution adheres to the approved ethics protocol. This careful process of technical engineering to satisfy research and governance requirements must be repeated for every study.

There Is a Significant Opportunity Cost for Creating Digital Solutions

Although individual research groups may achieve economies of scale through strategic development of shared code and the reuse of assets, the ability of other research groups to exploit these benefits is limited by the commitment of the originating researchers to make these assets available (eg, as open source repositories), to maintain them, and to provide support for their use. Researchers (especially those working outside electronic health [eHealth]) who may be interested in acquiring digital sensor signals may be essentially locked out because they have no access to enabling technology.

Efforts by mobile platform vendors to create reusable research tools (eg, Apple's ResearchKit [Apple Inc, Cupertino, California, USA]) do not address passive sensing scenarios and do not solve the problem of researchers wanting to build tools that will run on both iOS and Android platforms, since vendor frameworks are not cross-platform compatible.

Life Cycle Concerns Are Rarely Addressed

To guarantee participant experience, digital solutions must be supplied, set up, managed, and supported over the duration of the study. Significant effort and resources may be needed to address breaking changes introduced by updates to mobile operating systems and to debug issues experienced on specific device types (particularly for Android, where the device market is highly fragmented and where modifications to the device operating system by individual device manufacturers can create unexpected issues). The challenge of resourcing life cycle management is a potential contributor to the small numbers of research-backed health apps currently available in public app stores [28,36].

No Common Data Standards Exist for Digital Phenotyping

A data standard is a consensus set of rules for describing and recording information to facilitate its analysis, reuse, and exchange [37]. A topical challenge for digital phenotyping is the extent to which contextual information may be needed to adequately interpret signals data. For example, device type, version, and power state, sensor characteristics (such as measurement precision), and user characteristics (such as height and age) may all affect the interpretation of sensing data. Data standards offer a means to ensure that this relevant information is consistently collected in formats that are useful for both initial and secondary analyses. A lack of common standards acts as a potential barrier to combining data from multiple studies, with time and effort needed for data wrangling (ie, the process of converting data into compatible forms). Shared platforms stand to help address the standards gap in two ways. Firstly, simply by coordinating the activity of different researchers, they operate as *de facto* standards providers since all data collected is governed by the same technical collection mechanisms. Secondly, as the digital phenotyping landscape matures, they are well-placed to implement any standards that are collaboratively developed by the research community. Other areas where standardization may be beneficial include the specification of data preprocessing steps and the methods used to derive summary metrics and features intended for machine learning.

Good Data Governance is Challenging

Poor quality and privacy controls, including a lack of industry-standard safeguards such as appropriate encryption and access control, have repeatedly been identified in health apps available to the public [32,38]. Shared platforms may be better able to command the resources and expertise needed to monitor and respond to evolving governance risks than individual projects.

Design Considerations for Digital Phenotyping Platforms

In addition to factors motivating interest in coordinated approaches towards digital phenotyping, there are several specific design considerations. In a recent discussion [31], we identified three priorities. We argued firstly for the need for universal (ie, cross-platform) technology that is accessible regardless of device type, to ensure equity of opportunity for both research participants in the present and the potential future users of clinical services that are built around digital phenotyping. Despite almost equal market share between the two major mobile operating system vendors (Android and Apple) in multiple economies [39], digital phenotyping platforms have historically focused exclusively on Android devices because of the relative ease of implementing passive sensing on these compared to Apple, whose app model does not allow continuously running background services (except in specific circumstances not directly relevant to digital phenotyping).

Secondly, we advocated for platforms that could not only efficiently support smaller pilot and exploratory projects (of the kind that have largely characterized digital phenotyping research to date [31]) but also larger studies running with potentially thousands of concurrent participants. We justified this requirement on the basis that, if a primary intent of digital phenotyping is *de novo* biomarker discovery for clinical grade uses, then there will need to be a step change in the discriminative performance of models being derived from digital phenotyping signals, which will likely require larger datasets [40]. Reported sensitivities and specificities in digital phenotyping classification studies to date rarely exceed 90%, despite the use of state-of-the-art machine learning methods. We have highlighted how improvements in test statistics will be needed to obtain clinically acceptable rates of false positives and negatives [31].

Thirdly, and relatedly, we identified a need for platforms that could support the aggregation and efficient transformation of collected data to: (1) support its secondary reuse, such as the creation of artificial intelligence–based methods for predicting mental health outcomes; and (2) enable future integrations with digital health interventions. For example, models derived from digital phenotyping data could be used to tailor the selection and timing of components delivered by adaptive interventions. Platforms supporting this kind of operational use case have potentially more stringent requirements around uptime and resilience than those used simply as repositories for research data.

We identify here two further design considerations. The first is that the use of smartphones as the primary data collection mechanism introduces multiple constraints which necessitate trade-offs during platform design. That is, there are competing characteristics, such as user privacy concerns, energy efficiency for extending battery life, hardware difference due to device fragmentation, and strict app development guidelines [41-43]. Because each of these constraints has relevance for research-relevant goals, such as being able to collect as much

data as possible from a given user, there is a need for an explicit decision-making process during platform design.

Table 1 highlights the relationship between typical constraints and research goals, the tradeoffs that result, and potential design solutions. For example, there is an implicit trade-off between maximizing data collection volumes and potential negative impacts on user experience resulting from sensor-related battery drain and the use of data services to transmit collected data. For platform creators, design challenges include not only finding technical strategies that can optimize the efficiency of collection but also whether and how to include hard constraints that mitigate poor user experience (such as placing an upper limit on data collection frequency). Whether or not such strategies are necessary may depend, in turn, on the kind of operational model backing the platform. For example, where a platform is offered as a service to multiple research groups, the use of such constraints may be justified on the basis that negative user feedback arising from one study has the potential to affect the recruitment and participation of users in other studies, even if unrelated.

Table 1. Summary of the trade-offs between requirements, constraints, and resolution.

Research goal	Smartphone constraints	Trade-off	Potential design solution(s)
Readily accessible participant data	• Protect user's privacy by restricting access to data	• Data availability versus user privacy	• Strict adherence to mobile platform app guidelines captured in a reusable app development kit for app developers
Unified way to collect passive and active data	• Smartphone vendor fragmentation creating platform-specific data collection challenges, such as continuous background sensing on Apple devices	• High quality data versus Platform and device limitations	• Tested and verified implementation for accessing sensor data that wraps platform-specific data collection strategies in a common interface
High resolution data collection strategy (eg, continuous high frequency sampling over many days)	• Sensor usage limited to extend battery life • Impacts on network responsiveness and user data costs if cellular networks used to upload large data payloads	• High resolution data versus poor user experience • High resolution data versus additional costs to participant	• Custom communication protocol between mobile apps and core platform allows tailoring of frequency and duration of sensor sampling to manage energy and bandwidth impacts for users • Platform enforces upper limits on data collection resolution • Platform merges and schedules multiple requests to minimize impacts on users' devices • Support for customized scheduling of data uploads (eg, to use only Wi-Fi connectivity)

The second concerns the extent to which any platform enforces particular workflows on research users and data collection participants through its design. Even with modern software development approaches, assumptions about how users will perform key tasks are often incorporated at an early stage. As a result, there are limits to which the initial design can be modified [44], even with subsequent refactoring [45]. For example, if strong governance of collected data is a desirable outcome then it may be justifiable to introduce common platform mechanisms, such as requiring Institutional Review Board (IRB) approvals before data collection can be activated. However, our own experience as researchers also recommends against creating strong constraints on exactly what data research teams try to collect or the methods used for collection. One potential strategy for resolving this tension is to give study teams freedom regarding the design and function of client-side infrastructure (eg, data collection apps) while still mandating appropriate controls, including governance requirements, over server-side infrastructure where data are collected and processed for subsequent analysis. This approach has been successfully used in widely adopted software frameworks, such as those used to provide usage analytics on mobile apps.

The InSTIL (Intelligent Sensing to Inform and Learn) Digital Phenotyping Platform

Background

The InSTIL (Intelligent Sensing to Inform and Learn) platform is a new, cloud-based system for collecting active and passive sensor signals from both iPhone and Android smartphones.

Platform Design Aims

The platform was designed with the broad research aim of improving understanding of the causes and trajectory of youth-onset mood disorders using digital phenotyping. Requirements analysis was informed by the issues discussed above and by considering, through discussion with representative stakeholders, what would be needed to support multiple multidisciplinary teams working in parallel to explore different facets of this research challenge. These imagined teams consisted of: (1) mental health and clinical researchers wanting to explore relationships between specific, digital phenotyping–derived signals and traditional mental health outcomes, such as GPS data and self-rated depression scores, within observational studies of their own design; (2) researchers wanting to combine (with consent) collected datasets for secondary analyses, linkage, and machine learning; and (3) intervention designers wanting to consume digital phenotyping

data in some way to tailor or optimize new mental health interventions. It was assumed that teams would not necessarily be from the same institution and, although aligned with the overall research aim, would not necessarily focus on the same study populations.

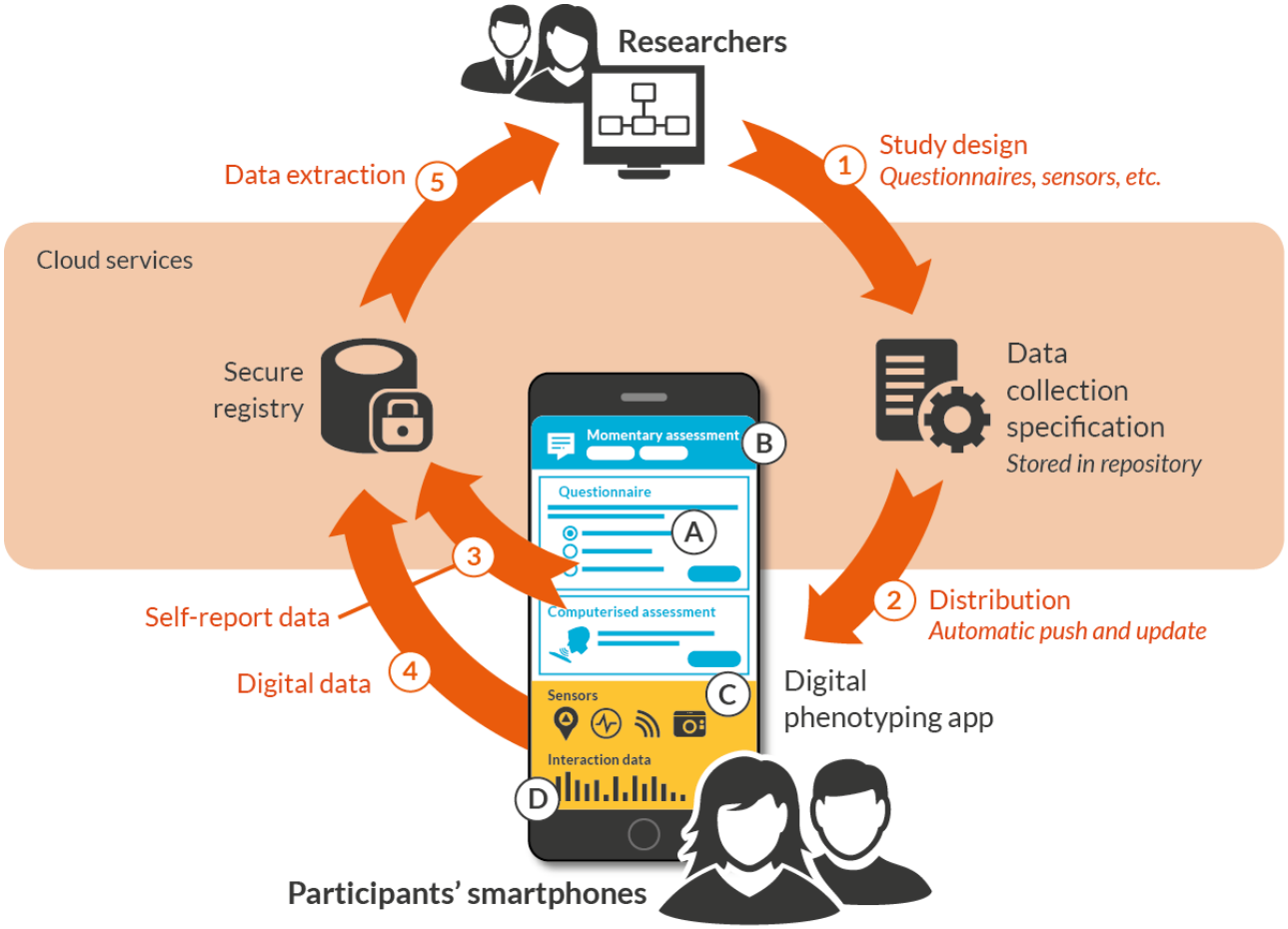
From this analysis, we identified three interrelated design objectives. Firstly, the platform should support high rates of reliable data ingestion (ie, the collection of high-resolution data from thousands of users without loss). This objective reflected both the identified need for, and research interest in, larger-scale digital phenotyping studies, as well as specific consideration of the data volumes required to establish a digital phenotyping bank. Platform design was influenced by this more than any other objective, reflecting the impact that performance and robustness requirements have on software architecture [46]. Secondly, the platform should be flexible enough to support the requirements of multiple studies, including those not based in mental health. In other words, the system needed to be flexible enough that researchers could adapt the platform to the specific requirements of their study while still benefiting from software reuse. Thirdly, the platform design should seek to minimize operational costs. Research teams often operate with constrained budgets with minimal allowance for the developer and operations teams that would be typical of such platforms in a commercial setting. This objective informed the ultimate decision to design certain platform components as shared backend services, thus removing the need for individual researchers to carry the risks and costs of initial setup and creating a potential mechanism for formal operational support in the future. Success in meeting this third objective is being assessed through cost modelling incorporated into a randomized controlled trial [47] that is using the platform for data collection. This will be reported in a future analysis.

Common Research Workflow

From the requirements gathering process we identified a common research workflow (Figure 1), which is a stereotyped sequence of actions for the acquisition of digital phenotyping data from study participants. This workflow incorporates the following steps:

1. Researchers specify the study design, define which questionnaires and sensors are required to deliver a trial or set of clinical measurements and, optionally, how these are integrated with any intervention components, such as self-guided therapy. This specification is then hosted in a secure online repository.
2. When the study commences, the specification is automatically distributed to users' devices. The same mechanism can be used to update the specification throughout the study.
3. The platform app development kit generates a range of data streams, as required. This includes self-report data streams, which consist of:
   - Self-report questionnaires and assessments. Self-report data can include single questions, standard instruments (such as the Patient Health Questionnaire [PHQ-9]), structured tests (such as response-time measurements) and recordings from sensors (such as voice samples); and
   - Ecological momentary assessments. These include quick-fire questions generated in response to temporal or contextual cues (for example, when a user awakes), to generate ecologically valid data.
4. There are also digital data streams, which might include:
   - Passive sensing, which runs silently and continuously to collect high resolution data about location, activity, and social interaction from the user. A contemporary smartphone may typically contain: an accelerometer, gyroscope, compass, barometer, light sensor, GPS receiver, microphone, camera, as well as Wi-Fi and Bluetooth interfaces that can be used to detect proximity to other users and devices. These sensor types are all routinely available for digital data collection.
   - Device utilization data, which tracks 'digital exhaust', such as time spent in apps, phone calls and text messaging usage, as well as potential markers of cognitive function, such as typing speed. The app development kit automatically manages potential barriers to data collection, such as user battery life and limited connectivity, through smart scheduling and caching.
5. Researchers can start to extract registry data as soon as it is received, accelerating analysis, permitting study designs that involve expert feedback, and allowing any data collection issues to be identified and addressed early in the research process.

Figure 1. InSTIL Common research workflow. The figure shows how the platform supports a sequence of key activities for researchers involved in digital phenotyping. InSTIL: Intelligent Sensing to Inform and Learn.



The workflow makes explicit several user requirements that informed platform design. Firstly, and reflecting the status of digital phenotyping as an emerging research area, an iterative "test, analyze, test" approach is likely to be the norm, with the timing and focus of digital phenotyping data collection revised through a sequence of sub-studies. This requirement drives the cyclical graph shown in Figure 1. Any iterative approach to data collection must nevertheless be balanced with good study governance requirements such as data integrity and versioning.

Secondly, researchers who are faced with practical study issues, such as unforeseen protocol adjustments or managing participant compliance, will expect the flexibility to amend data collection parameters on demand. This gives rise to the concept of a dynamic, study-specific, data collection specification that can be updated and disseminated to participant devices as required.

Thirdly, support for acquisition of nondigital phenotyping data types is essential to enable key uses cases, such as the need for self-report and outcomes measures to provide labels against which to train digital phenotyping models using supervised machine learning.

Finally, despite the need for study-specific software development work (eg, engineering apps to satisfy specific data collection, intervention, and user experience requirements), there is still value in common, researcher-facing software tools to support routine administrative tasks such as scheduling, monitoring participation, and data extraction. Relatedly, the ways in which researchers will use collected data, whether for statistical analysis, machine learning, or intervention tailoring, are diverse enough to limit the value of single mechanisms for visualizing and manipulating data. Instead the focus should be on providing ways to efficiently extract collected data in useful formats and slices (ie, subsets of users).

Platform Components

The resulting platform consists of a set of reusable software components embedded within a common architecture (see Figure 2 and summary in Table 2). The architecture is consistent with identified requirements for big data platforms, specifically those of distributed computation and big data storage [48]. Together, these components: (1) enable the collection of passive and active sensor data (and other arbitrary data types); (2) enable secure storage of deidentified data; (3) provide dynamic control over the frequency and types of data being collected; (4) maintain the provenance of data by ensuring source information is recorded consistently; (5) provide a robust synchronization protocol to reduce the risk of data corruption during transfer; (6) record data in a standardized and stable data format to permit replication and auditing; and (7) offer a common data export and management method. The components shown in Table 2 have been designed to adhere to the software engineering principle of separation of concerns [49], where each component focuses on a well-defined functional process to minimize repetition of code, maximize reusability, and simplify maintenance or further development.

Figure 2. Platform architecture. Architecture diagram showing the digital phenotyping platform consisting of reusable components (purple and blue boxes) and components that need to be custom built for each study (orange).

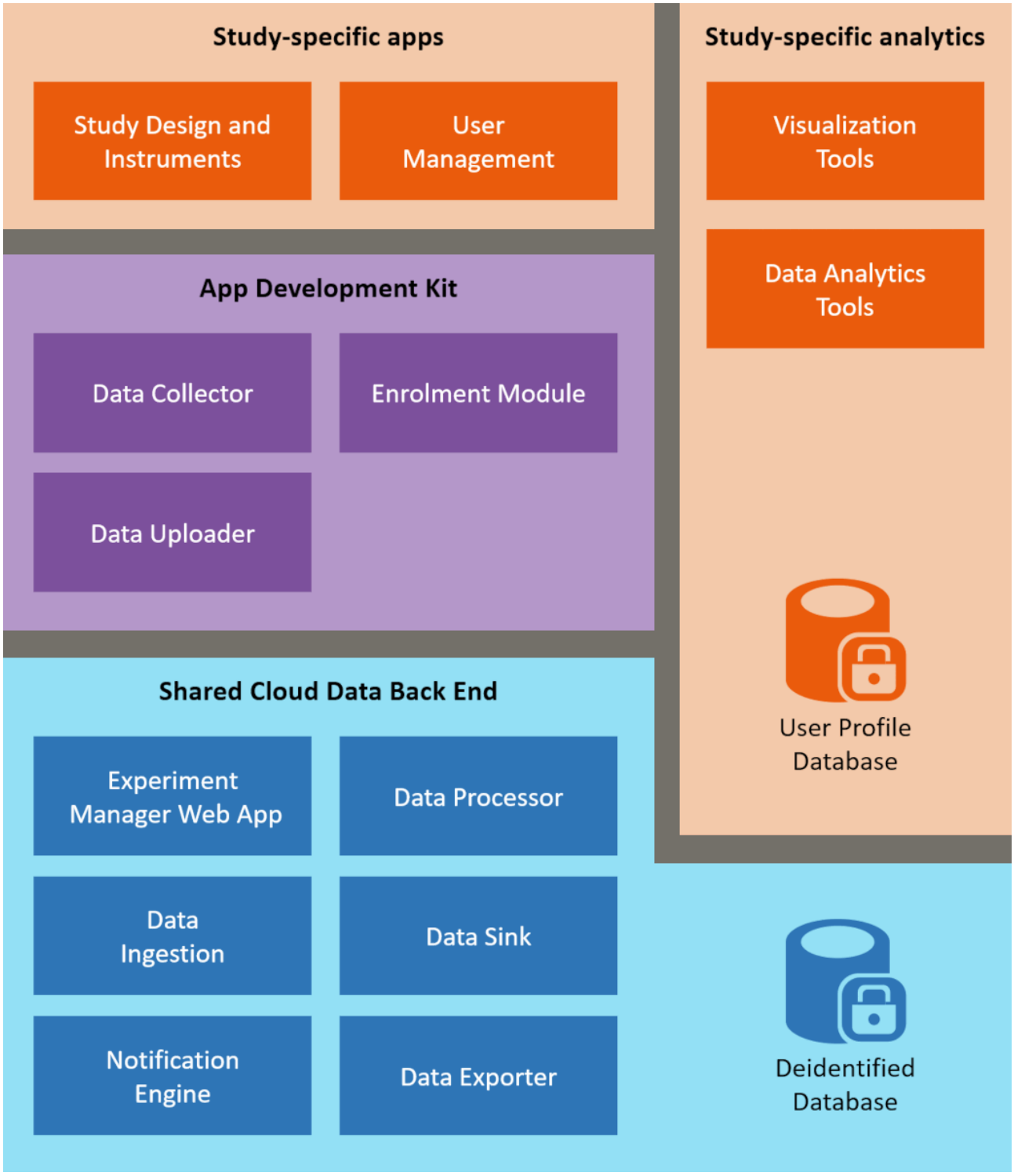


Table 2. Description of architecture components.

Component	Description
Study design and instruments	Questionnaire, data collection strategy, and the frequency required for a specific experiment (active data).
User management	Some studies may need to reidentify users or provide custom authentication to ensure that only specific participants join the study, and this component enables custom apps to provide that functionality.
Visualization tools	Integration endpoint for third party data visualization tools (eg, Power BI, Tableau, or Quicksight).
Data analytics tools	Integration endpoint for big data tools to provide analytics and support machine learning techniques.
User profile database	Database for storing study specific details. This store is independent of the platform, thus preserving the separation between identifiable data (held in this database) and anonymized data (held in cloud data stores).
Data collector	Component responsible for extracting passive data from the mobile device and storing it locally prior to uploading the data to the platform.
Data uploader	Component responsible for uploading data to the cloud backend. Fault tolerance and automatic resume-retry supported.
Enrolment module	Responsible for enrolling a participant and device with a specific experiment.
Experiment manager web app	Dedicated component responsible for creating a new experiment with all the details specific to the study.
Data ingestion	Accepts raw data from the mobile devices and triggers the data processor to store all data for the user.
Data processor	Backend component responsible for transforming the data to the Data Sink component.
Data sink	Component responsible for storing the study data in a database.
Notification engine	Triggers notifications to be sent to the mobile apps to ensure that they continue to collect the appropriate data required for the study.
Data exporter	Module that researchers can use to extract data from the database in a standardized format.
Deidentified database	Core database that stores all deidentified data collected from an app.

Platform components are divided between a mobile app development kit and shared cloud-based data backend. The app development kit is available for both iOS and Android and is intended to expedite development of new digital phenotyping client apps. The app development kit reflects our skepticism that any fully-fledged data collection app can meet the design, content, and user experience requirements of any given study. Instead, native libraries included in the kit provide drop-in passive and active sensor data collection capabilities (summarized in Figure 3), and separate error-tolerant and secure upload management for collected data. These libraries allow any app to be augmented with digital phenotyping capabilities, while leaving researchers full control over the user-facing interface and interaction design. The app development kit also allows researchers to exert fine control over energy and data impacts associated with sensor utilization and data upload.

The common data platform provides endpoints to authorize, ingest and securely store data uploaded by apps using the app development kit, allows researchers to dynamically configure the behavior of data collection apps, makes collected data available to researchers for further analysis, and maintains persistent audit logs of data upload and access for governance purposes. Although individual studies can set up and run their own instance of the common data platform, it is principally designed to support the use case where it can be hosted and offered as a service to multiple concurrent projects. The design is intended to reduce the learning curve, costs, and risks borne by users new to the platform who can instead focus on the client-side data collection experience. In our solution, researchers interact with the platform through Experiment Management, a web app (interface shown in Figure 4) which allows them to specify data collection configurations, monitor data collection progress, and download data collected from previous experiments. The platform supports the creation of custom study designs (questionnaires, interactions, etc) and flexible data collection protocols (nightly uploads, type and frequency of sensor data, etc). It also acts as a formal store for audit-relevant information about study permissions, such as documents received from an authorizing IRB. Without this information, the study cannot commence data collection.

Figure 3. Supported data collection types. EMA: ecological momentary assessment.

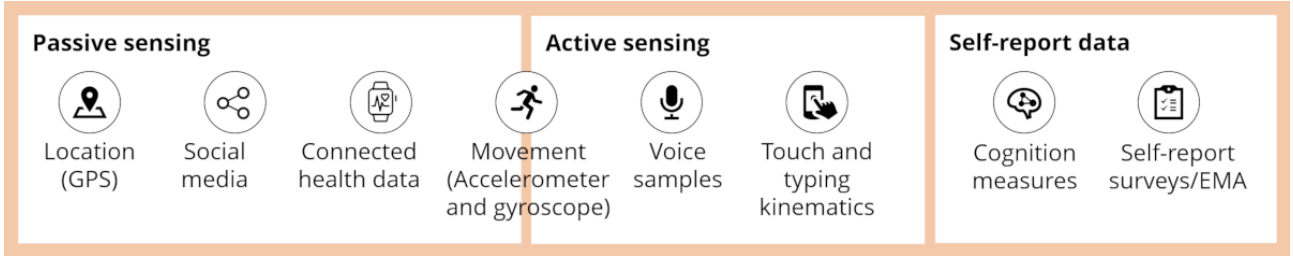
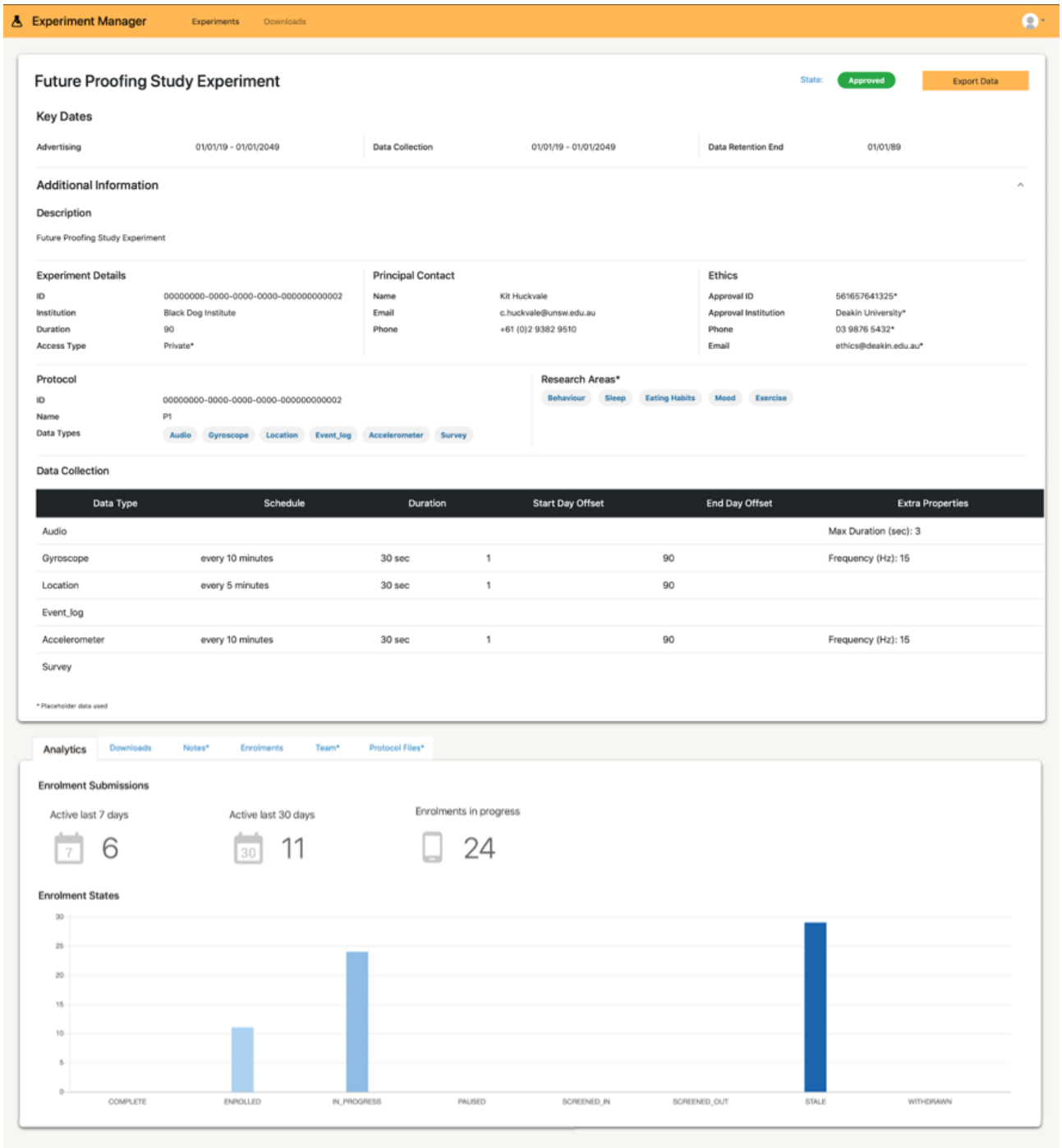


Figure 4. Experiment manager interface. Screenshot of the Experiment Manager application showing a summary of a study, including the data being collected (top half), and the enrollment status for the participants (bottom half).



Platform Features

Support for Passive Sensing on Both iOS and Android Devices

The platform app development kit provides native libraries to support app development on the devices of both major platform vendors. The ability to schedule passive sensing on Apple devices differentiates ours from most other digital phenotyping platforms. The provision of a kit ensures that researchers do not need to reimplement best practices associated with privacy and security in a mobile cloud computing environment, such as appropriate authorization and encryption during data upload [50].

Strategies to Maximize Passive Sensing Completeness

A principal challenge for passive sensing on smartphones is to maximize the proportion of desired data points that are sampled as planned and successfully returned for analysis. Although data upload scheduling and error tolerance is an important modifier of data completeness, a major barrier concerns reliable sampling (ie, the initial acquisition of data from the desired sensor at the desired time).

Concerns around data privacy have led mobile platform vendors to introduce restrictions designed to give users greater control

over how sensor data are collected and used. For example, mobile apps must now explicitly obtain (and in some cases periodically reobtain) user permission to collect GPS location data. These permission mechanisms are platform-specific and subject to change. To address these challenges, our development kit provides support for a flexible onboarding mechanism to guide users through the process of setting appropriate permissions, configuring whether cellular data can be used for data upload, and configuring hooks for notifying where permissions have been denied or revoked. Fault tolerance means that the kit handles such cases gracefully, continuing to collect other sensor types even if one sensor no longer has permission.

Another limitation arises from concerns about the potential impact of continuous background sensing on system performance and battery drain adversely affecting the end-user experience [43]. Platform vendors limit the amount of data that can be collected passively, even closing apps that are energy inefficient or have not been recently activated by the user. To work within the constraints imposed by the platform vendors, our communication protocol relies on periodically waking up the mobile apps using silent background notifications. This approach allows a common mechanism to be used on both Android and iOS but is reliant on network connectivity. The efficacy of these strategies in maximizing data completeness is the subject of current investigation and we intend to report on it in a future manuscript.

Extensible, Low-Cost Data Ingestion Pipeline

Data management has shown to be a core consideration in large-scale, mobile app ecosystems [51]. This principle influenced the decision to make the data ingestion pipeline hosted by the cloud platform essentially agnostic to the form and type of data uploaded from users' devices. Instead, uploads are required to be decorated with metadata, which encodes information type and shared parameters, such as time of collection, in a standardized format. This flexibility means that support for new sensor types or data sources, or changes to the form in which data are collected, can easily be introduced to data collection clients without impacting the function of the cloud-based ingestion pipeline. For example, a research group wanting to extend the app development kit to collect high volume data from a wearable sensor could do so and continue to use existing data upload mechanisms without requiring changes to the cloud architecture. This flexibility is intended to expedite this kind of development and reduce the associated costs for both research users and platform operators. The use of standardized metadata means that analytics functions, such as data submission rates, can easily be derived from submitted data even if new types have been added.

In addition to extensibility, the data ingestion pipeline is designed to minimize overheads associated with data acquisition by performing minimal computation over the ingested data. For example, data validation and compression are not performed by the pipeline but are, instead, a client-side concern supported by the app development kit. By distributing computation to client devices, this reduces potential bottlenecks and consequent costs for the common platform.

Strongly Typed, Client-Side Data Collection

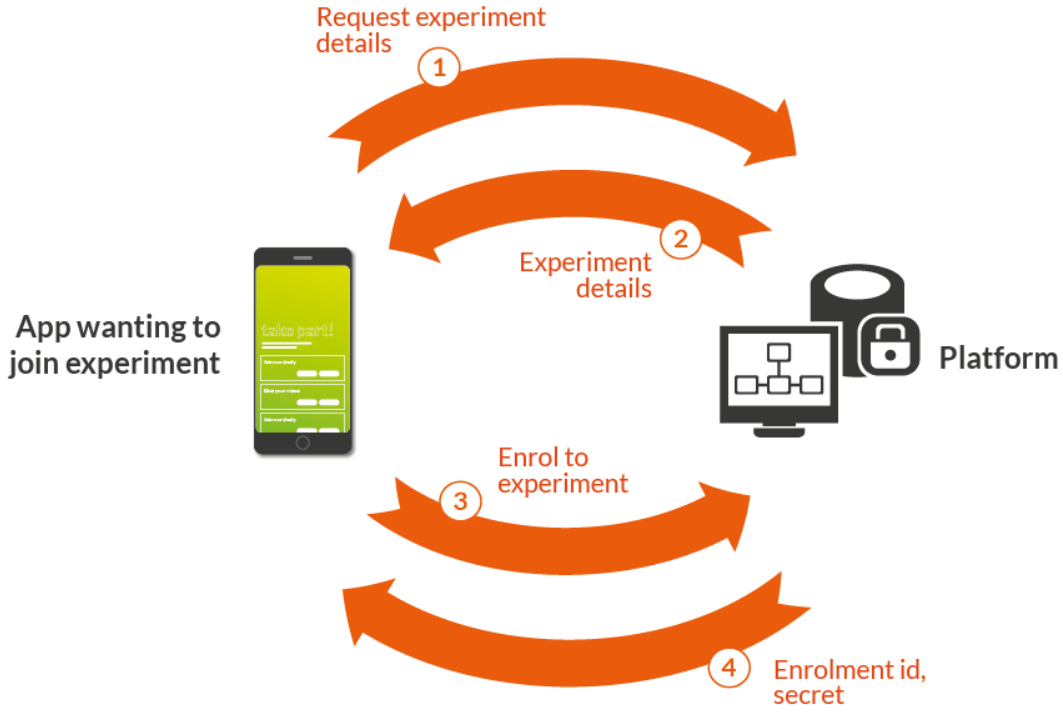
While the data ingestion pipeline is type-agnostic, the app development kit implements a core set of strongly typed schemas for capturing and persisting common sensor data and related data types. Schemas are currently available for GPS, accelerometry, gyroscope, audio, longitudinal event log, and questionnaire data and can be freely extended by users of the kit. This combination of an agnostic pipeline with client-side conventions for storing data represents a design trade-off between the flexibility to extend and adapt the platform and the ability to reuse, transform, and combine data (which recommends the use of common data standards). As a result, research users of the platform that implement the development kit can use the provided data types and be reassured that the data they collect will be compatible with existing datasets as well any visualization/analysis tooling developed for the platform, while those who want to adapt the data formats can do so freely.

Privacy-Preserving Architecture

Consumer privacy law and research ethical guidelines create strict requirements about data handling and management [52,53]. Considerations such as data privacy, which jurisdiction the data can be stored and transferred to, and system security are all relevant governance concerns addressed by the platform. For example, the selected cloud/provider architecture allows us to assure (with suitable configuration) that data belonging to each research study are appropriately segregated and will be stored only in a single legal jurisdiction, while the common data platform enforces a requirement for IRB approvals to be lodged prior to starting data collection.

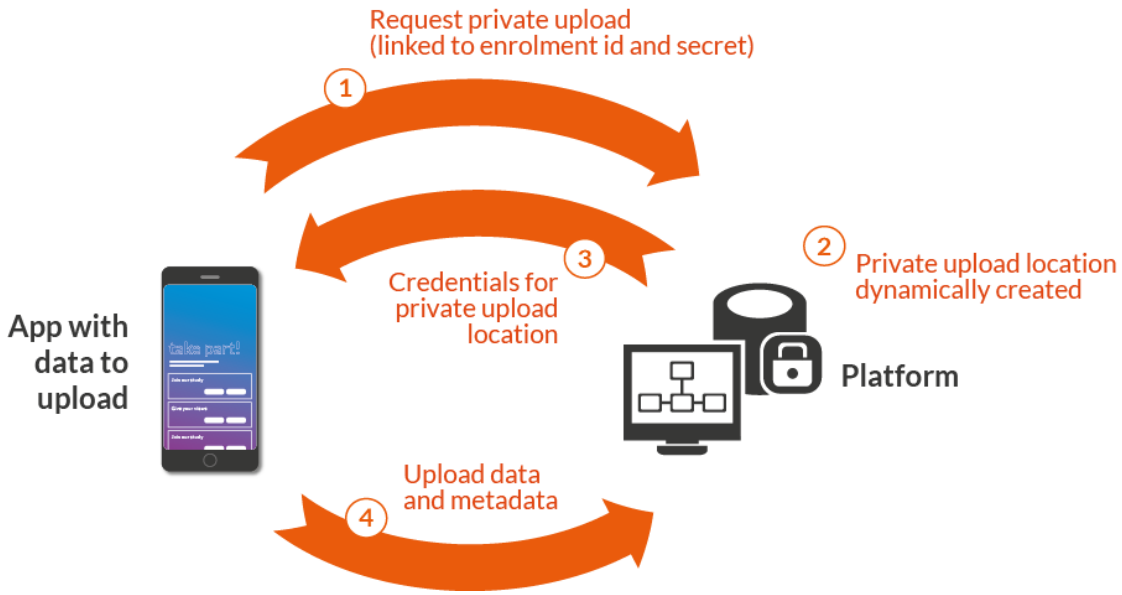
A distinguishing feature of the platform is that all user data contributions are anonymous by default. This is achieved using a custom anonymous authentication mechanism (see Figure 5) which is provided as part of the app development kit and enforces contractual anonymity within the mobile app. We recognize, however, that for certain studies it may be necessary to identify and follow up with named individuals to augment or verify results or deliver other intervention components. The platform supports the case where participants do need to be identified, but this requires explicit additional steps to be taken by developers (in addition to appropriate ethical permissions) to cache the unique enrollment identifier generated for every participant. To do this, developers must create their own linkage key store. This is, by definition, separate from the data stores used by the platform, preserving anonymity in the unlikely event that the platform stores are compromised.

Figure 5. Privacy enforced enrolment protocol. Protocol for communication between the digital phenotyping app and the cloud backend supporting passwordless authentication.



A second feature is the use of public key-based mechanisms as a strategy to authorize and secure the time-limited transfer of user data payloads to the platform. This mechanism (summarized in Figure 6) guarantees that uploaded data are uniquely allocated to the user that initiated the upload, minimizing the potential for unauthorized access, upload spam, or mistakes in data allocation during analysis. Together, these strategies mean that there is no situation during routine platform operation in which identifiable information is stored in the same cloud environment as user response data.

Figure 6. Secure data upload protocol. Protocol for communication between the digital phenotyping app and the cloud backend involving anonymous data collection and dynamically created upload locations.



Technical Implementation Overview

A central tenet of the technical implementation was to leverage existing open source components wherever possible. Due to the scalable requirements of the platform, which is the eventual ability to support tens of thousands of concurrent users, the platform was built on top of the containerized platform Kubernetes (Google LLC, Mountain View, California, USA). Selecting a containerized solution simplified the processes of integrating existing tools into the platform. These included Elasticsearch (Elastic NV, Amsterdam, the Netherlands) for efficient information retrieval, Kafka (Apache Software Foundation, Forest Hill, Maryland, USA) for constructing data pipelines, Zookeeper (Apache Software Foundation, Forest Hill, Maryland, USA) for coordinating distributed components, Kibana (Elastic NV, Amsterdam, the Netherlands) for visualization and reporting, Redis (Redis Labs, Mountain View, California, USA) as a central data store and PostgreSQL (PostgreSQL Global Development Group, Berkeley, California, USA) for data persistence. Scripts are provided to configure, deploy, and run these components on Google's Cloud Platform (Google LLC, Mountain View, California, USA) (although the platform will run on any cloud environment that supports Kubernetes). The applications that make up the cloud backend are written as Java microservices using the Spring Boot (Pivotal Software, San Francisco, California, USA) web framework. Digital phenotyping apps use platform specific native programming languages, such as Swift for Apple's iOS and Java/Kotlin for Google's Android.

Future Development

A principal task for future development is to identify where shared value can be created for users once data have been received by the platform. While we expect that most researchers will expect direct access to raw data, there may be tasks, particularly those that are either computationally expensive or technically sophisticated, which it would make sense to be offer as platform services. For example, feature computation or data reduction techniques could be offered as options within a standardized processing pipeline, minimizing set-up burden for platform users while offering potential assurance about the standardization of the datasets that are produced. Work is underway to explore the scope, feasibility, and value of this kind of preprocessing.

Additional work is also required to enhance error handling logic for the platform. To achieve high data throughput, validation logic (ie, steps to verify the correctness of incoming data) within the ingestion pipeline was kept to a minimum. While this pipeline can theoretically support efficient validation, such as through asynchronous workers operating over accumulated data, the current architecture does not support mechanisms to propagate identified validation errors (eg, poor sample quality) back to mobile clients. This introduces potential latency into the process of identifying data quality issues and shifts the burden of problem-solving onto the research team as a manual process, which may be unfeasible in studies with large numbers of participants. Future updates will introduce mechanisms to trigger app development kit hooks when validation issues are discovered, providing a potential route for data collection clients to respond and fix any identified validation problems without manual intervention. For example, in a study collecting active voice samples, a collected sample that fails a quality check could trigger the platform-associated mobile app to ask a participant to provide a second sample.

We also recognize an ongoing requirement to review and strengthen approaches to safeguarding participant confidentiality. While the platform design ensures that traditional identifiers, such as names and contact details, are not collected, we acknowledge the risk that certain digital phenotyping data types, such as GPS, are intrinsically identifiable [54]. Potential platform-enforceable strategies to manage this risk include: holding data in temporal escrow to prevent individuals' current location being identified, remapping GPS data to an alternate reference frame so that relationships between points, but not their absolute location, are preserved, and providing pipeline tools to preprocess data into features or labels on behalf of researchers to avoid the need for teams to handle raw data. Each of these involves potential trade-offs between risk management and analytical value [55]. For example, remapping GPS to an arbitrary coordinate system limits the potential for semantic labelling of known locations. We aim to explore the feasibility and acceptability of these kinds of strategies in relation to both GPS and other sensor data types in the next phase of work.

Finally, a process of technical refinement (currently underway) aims to simplify the existing architecture to: (1) eliminate unnecessary components which nevertheless confer operational costs; and (2) reduce the effort required to extend the platform for study-specific modifications. Currently, binary data such as audio, video, and image files are stored in Elasticsearch, enabling all data to be indexed for efficient searching. In later development phases, blob storage will be used to house binary data and only metadata describing the files will be indexed, with the aim of reducing data processing costs. This shift may also make it possible to replace Elasticsearch with a simpler data store, such as a relational database, potentially reducing the skills required of the team operating the platform.

Discussion

In this paper we describe the design of a scalable and governance-aware platform for the acquisition of sensor data from consumer smartphones, for the purposes of digital phenotyping. The InSTIL platform provides a new suite of tools for the development of digital phenotyping research studies. The platform is currently being used to run studies at Black Dog Institute and Deakin University, and will ultimately be made available for research and public use. It sits alongside several established and emerging technology platforms for digital phenotyping that include Aware [56], Beiwe [57,58], EARS [59], Purple [60], Monsenso [61], Passive Data Kit [62], and RADAR-base [63]. InSTIL differs from these existing platforms in several ways: (1) it supports passive sensing on iOS, unlike EARS, Purple, and RADAR-base; (2) it runs sensing without using special permissions that may prevent apps from being deployed via public app stores, unlike Aware (albeit at the expense of certain sensor streams, such as Bluetooth proximity

tracking); (3) and its hosted server model means that research teams do not have to worry about setting up, securing, and maintaining the server infrastructure needed to run digital phenotyping studies, unlike Passive Data Kit. Our platform appears to be conceptually like Beiwe as a service [64], adopting similar positions on issues such as participant anonymization [65]; however, InSTIL differs by pairing its hosting model with a software development kit, enabling custom apps to be built for each project rather than offering standard data collection apps (as Beiwe does). Each approach has pros and cons concerning the ability to tailor user experience versus development cost and timelines.

InSTIL is also an enabler for a research vision that seeks to establish an international digital phenotype bank to pool sensor signals data from a population-scale longitudinal cohort. This initiative creates the opportunity to link digital phenotyping data to ground truth data, including access to DNA information, hospital records, and educational outcomes, to sensor data within individual trials, projects, and experiments. So far, platform development has focused primarily on technical aspects of data collection, including the requirement to support the volume of incoming data that would be necessary for a population scale bank (potentially millions of data samples, per participant, per year). The data collection platform is not the bank, however, and we recognize that substantial additional work will be required to address technical, governance, and operational issues around efficient and secure data sharing (particularly across international boundaries), participant recruitment, and management.

From a technical perspective, there are several insights arising from the development process which may be relevant to other researchers involved in similar digital phenotyping or other platform-scale efforts. Firstly, for managing even high-volume data ingestion scenarios, a simple publish-subscribe messaging queue may suffice. In retrospect, our selection of Kafka as the queue management technology, while not creating any active roadblocks, now provides few benefits. While Kafka can perform persistent queries on the queue itself and replay data streams, these features add limited value in the final system. Second, the initial idea of simplifying the architecture by using a comprehensive search engine (Elasticsearch) for all different types of data proved to be unnecessary, as the principal data storage format ended up being JSON. A simple document storage database or a relational database system with JSON indexes would be enough. Even where binary data are being collected, such as audio samples, there is no specific use case for holding these in an indexable store. Rather, it is the outputs from analysis of these data packets (eg, machine learning/algorithm-derived labels and annotations) where indexing may be helpful, and since we propose that these outputs be in JSON format, these too can be kept in a relational or document store. Together these issues highlight the challenge of accurately forecasting appropriate technology selection in complex projects, particularly those developed using iterative approaches. Finally, building and maintaining a digital phenotyping platform requires a thorough understanding of the limitations and constraints put in place by the mobile platform vendors. This is a challenging task as mobile operating systems are frequently updated, often with breaking changes. For example, changes to the latest version of iOS significantly reduced the ease with which passive location data could be continuously collected.

In this paper, we also highlighted the trade-offs inherent in designing data collection infrastructure that runs on users' own devices, which researchers have limited control over and where inconvenience must be minimized. In addition to these user-facing tradeoffs, the overall platform architecture itself represents a trade-off between the desire to give researchers maximum flexibility in configuring data collection to suit their research and user experience requirements, while minimizing the costs of completing technically challenging tasks such as passive sensing and a secure data upload. We use the client-server divide to orchestrate this trade-off, with researchers given freedom on the app (client) side while being supported by a common (server) backend with minimal scope for study-specific changes. Work is now needed to critically evaluate the value of this strategy, particularly since we have already seen evidence of how it might be challenged. While developing the app development kit, the team discovered that the iOS App Store review process prevents apps from collecting location data passively without surfacing this information in a feature that directly benefits the end user. A design question arises, then, as to whether the platform should provide native visualization of GPS location data in some user-facing format (which may require technical sophistication to achieve), or whether this should be a study-specific concern. As the platform is designed to be useful across many different studies, implementation of such a feature may be beneficial but comes with development and maintenance costs. It seems likely that, should the platform become more widely used, a prioritization mechanism will be needed to select features for inclusion in the development kit/backend. We also recognize that the putative benefits of a common and shared infrastructure are largely linked to the ability to provide operational support to projects. As a result, the next phase of platform development will also focus on the feasibility of different operational and governance models.

Although in the early stages of use, our vision for this new science of behavior [2] promises benefits for health care, digital health industries, and science. For health care, it provides the means for users to learn more about their mental health, to be able to, once sufficient work has been completed, anticipate when their health may be at risk, and be able to access techniques and interventions that have proved successful for themselves and for many others. Ultimately, this work is aimed at optimizing health care and alleviating disempowerment and emotional toil. For the fast-growing digital industry, it raises the possibility of new commercial opportunities, such as tools for health insurers that incentivize health behaviors through passive monitoring. For scientists, a collaborative platform to generate, combine, and use signals reduces the opportunity costs of digital phenotyping research and opens the door to new multidisciplinary collaborations.

There is some understandable skepticism about the potential of digital sensor signals to provide meaningful data to help manage mental illness. Our own uncertainties have been reduced, in part

by the success of modern machine learning methods in tackling previously intractable classification problems in multiple domains, by promising results from early digital phenotyping studies [31], and by evidence supporting the existence of prodromal phases with behavioral correlates for common mental illnesses [66,67]. If these changes can manifest in sensor data streams, then they can be detected and modelled. Ultimately, we will not know if this can be a cost-effective, acceptable, and robust approach unless we try. A collaborative digital phenotyping platform, open to multiple users working in parallel, will be critical to answer this question quickly.

Authors' Contributions

KH developed the initial concept for the platform and solicited/collated user requirements. RV and SB designed the system architecture. SB wrote, and RV critically reviewed, the first draft of this manuscript. KH wrote the second draft with input from HC. All authors reviewed and provided comments. All authors had access to and approved the final version of the manuscript.

Conflicts of Interest

HC is the director of Black Dog Institute, which develops apps and internet interventions for mental health, but has no personal financial gain. HC stands to receive royalties as a creator of MoodGYM, but to date has no financial gain. All authors are involved in the development of the InSTIL platform described in this manuscript.

References

1. Hollis C, Morriss R, Martin J, Amani S, Cotton R, Denis M, et al. Technological innovations in mental healthcare: harnessing the digital revolution. Br J Psychiatry 2015 Apr;206(4):263-265. [doi: 10.1192/bjp.bp.113.142612] [Medline: 25833865]
2. Insel TR. Digital Phenotyping: Technology for a New Science of Behavior. JAMA 2017 Oct 03;318(13):1215-1216. [doi: 10.1001/jama.2017.11295] [Medline: 28973224]
3. Onnela J, Rauch SL. Harnessing Smartphone-Based Digital Phenotyping to Enhance Behavioral and Mental Health. Neuropsychopharmacology 2016 Jun 28;41(7):1691-1696 [FREE Full text] [doi: 10.1038/npp.2016.7] [Medline: 26818126]
4. Mohr DC, Zhang M, Schueller SM. Personal Sensing: Understanding Mental Health Using Ubiquitous Sensors and Machine Learning. Annu Rev Clin Psychol 2017 May 08;13(1):23-47. [doi: 10.1146/annurev-clinpsy-032816-044949] [Medline: 28375728]
5. Wahle F, Kowatsch T, Fleisch E, Rufer M, Weidt S. Mobile Sensing and Support for People With Depression: A Pilot Trial in the Wild. JMIR Mhealth Uhealth 2016 Sep 21;4(3):e111 [FREE Full text] [doi: 10.2196/mhealth.5960] [Medline: 27655245]
6. Mundt JC, Vogel AP, Feltner DE, Lenderking WR. Vocal acoustic biomarkers of depression severity and treatment response. Biol Psychiatry 2012 Oct 01;72(7):580-587 [FREE Full text] [doi: 10.1016/j.biopsych.2012.03.015] [Medline: 22541039]
7. Place S, Blanch-Hartigan D, Rubin C, Gorrostieta C, Mead C, Kane J, et al. Behavioral Indicators on a Mobile Sensing Platform Predict Clinically Validated Psychiatric Symptoms of Mood and Anxiety Disorders. J Med Internet Res 2017 Mar 16;19(3):e75 [FREE Full text] [doi: 10.2196/jmir.6678] [Medline: 28302595]
8. Stange J, Zulueta J, Langenecker S, Ryan K, Piscitello A, Ajilore O. Convergence of active and passive assessments of affective instability in predicting the prospective course of bipolar disorder: The bi affect study. Neuropsychopharmacology 2017 Nov;42(Supplement 1):S111-S293 [FREE Full text] [doi: 10.1038/npp.2017.264]
9. Zulueta J, Andrea P, Rasic M, Easter R, Babu P, Langenecker S, et al. 481. Predicting Mood Disturbance Severity in Bipolar Subjects with Mobile Phone Keystroke Dynamics and Metadata. Biological Psychiatry 2017 May;81(10):S196. [doi: 10.1016/j.biopsych.2017.02.965]
10. Kessing LV, Munkholm K, Faurholt-Jepsen M, Miskowiak KW, Nielsen LB, Frikke-Schmidt R, et al. The Bipolar Illness Onset study: research protocol for the BIO cohort study. BMJ Open 2017 Jun 23;7(6):e015462 [FREE Full text] [doi: 10.1136/bmjopen-2016-015462] [Medline: 28645967]
11. Faurholt-Jepsen M, Frost M, Martiny K, Tuxen N, Rosenberg N, Busk J, et al. Reducing the rate and duration of Re-ADMISsions among patients with unipolar disorder and bipolar disorder using smartphone-based monitoring and treatment - the RADMIS trials: study protocol for two randomized controlled trials. Trials 2017 Jun 15;18(1):277 [FREE Full text] [doi: 10.1186/s13063-017-2015-3] [Medline: 28619114]
12. Gruenerbl A, Osmani V, Bahle G, Carrasco J, Oehler S, Mayora O. Using smart phone mobility traces for the diagnosis of depressivemanic episodes in bipolar patients. 2014 Presented at: 5th Augmented Human International Conference; 7-8 Mar; Kobe, Japan URL: https://dl.acm.org/citation.cfm?doid=2582051.2582089 [doi: 10.1145/2582051.2582089]
13. Palmius N, Tsanas A, Saunders KEA, Bilderbeck AC, Geddes JR, Goodwin GM, et al. Detecting Bipolar Depression From Geographic Location Data. IEEE Trans Biomed Eng 2017 Aug;64(8):1761-1771 [FREE Full text] [doi: 10.1109/TBME.2016.2611862] [Medline: 28113247]
14. Raballo A. Digital phenotyping: an overarching framework to capture our extended mental states. The Lancet Psychiatry 2018 Mar;5(3):194-195 [FREE Full text] [doi: 10.1016/s2215-0366(18)30054-3]

15. Fitzpatrick KK, Darcy A, Vierhile M. Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial. JMIR Ment Health 2017 Jun 06;4(2):e19 [FREE Full text] [doi: 10.2196/mental.7785] [Medline: 28588005]
16. Vaidyam AN, Wisniewski H, Halamka JD, Kashavan MS, Torous JB. Chatbots and Conversational Agents in Mental Health: A Review of the Psychiatric Landscape. Can J Psychiatry 2019 Jul;64(7):456-464. [doi: 10.1177/0706743719828977] [Medline: 30897957]
17. Arean PA, Hallgren KA, Jordan JT, Gazzaley A, Atkins DC, Heagerty PJ, et al. The Use and Effectiveness of Mobile Apps for Depression: Results From a Fully Remote Clinical Trial. J Med Internet Res 2016 Dec 20;18(12):e330 [FREE Full text] [doi: 10.2196/jmir.6482] [Medline: 27998876]
18. Witt K, Spittal MJ, Carter G, Pirkis J, Hetrick S, Currier D, et al. Effectiveness of online and mobile telephone applications ('apps') for the self-management of suicidal ideation and self-harm: a systematic review and meta-analysis. BMC Psychiatry 2017 Aug 15;17(1):297 [FREE Full text] [doi: 10.1186/s12888-017-1458-0] [Medline: 28810841]
19. Nicholas J, Larsen ME, Proudfoot J, Christensen H. Mobile Apps for Bipolar Disorder: A Systematic Review of Features and Content Quality. J Med Internet Res 2015 Aug 17;17(8):e198 [FREE Full text] [doi: 10.2196/jmir.4581] [Medline: 26283290]
20. Ruwaard J, Lange A, Schrieken B, Emmelkamp P. Efficacy and effectiveness of online cognitive behavioral treatment: a decade of interapy research. Stud Health Technol Inform 2011;167:9-14. [Medline: 21685634]
21. Ahern E, Kinsella S, Semkovska M. Clinical efficacy and economic evaluation of online cognitive behavioral therapy for major depressive disorder: a systematic review and meta-analysis. Expert Rev Pharmacoecon Outcomes Res 2018 Feb;18(1):25-41. [doi: 10.1080/14737167.2018.1407245] [Medline: 29145746]
22. Firth J, Torous J, Nicholas J, Carney R, Pratap A, Rosenbaum S, et al. The efficacy of smartphone-based mental health interventions for depressive symptoms: a meta-analysis of randomized controlled trials. World Psychiatry 2017 Oct;16(3):287-298 [FREE Full text] [doi: 10.1002/wps.20472] [Medline: 28941113]
23. Firth J, Torous J, Nicholas J, Carney R, Rosenbaum S, Sarris J. Can smartphone mental health interventions reduce symptoms of anxiety? A meta-analysis of randomized controlled trials. J Affect Disord 2017 Aug 15;218:15-22 [FREE Full text] [doi: 10.1016/j.jad.2017.04.046] [Medline: 28456072]
24. Linardon J, Cuijpers P, Carlbring P, Messer M, Fuller-Tyszkiewicz M. The efficacy of app-supported smartphone interventions for mental health problems: a meta-analysis of randomized controlled trials. World Psychiatry 2019 Oct 09;18(3):325-336 [FREE Full text] [doi: 10.1002/wps.20673] [Medline: 31496095]
25. Holmes EA, Ghaderi A, Harmer CJ, Ramchandani PG, Cuijpers P, Morrison AP, et al. The Lancet Psychiatry Commission on psychological treatments research in tomorrow's science. Lancet Psychiatry 2018 Mar;5(3):237-286. [doi: 10.1016/S2215-0366(17)30513-8] [Medline: 29482764]
26. Christensen H, Mackinnon A. The law of attrition revisited. J Med Internet Res 2006 Sep 29;8(3):e20; author reply e21 [FREE Full text] [doi: 10.2196/jmir.8.3.e20] [Medline: 17032636]
27. Torous J, Nicholas J, Larsen ME, Firth J, Christensen H. Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements. Evid Based Ment Health 2018 Aug 05;21(3):116-119. [doi: 10.1136/eb-2018-102891] [Medline: 29871870]
28. Larsen ME, Huckvale K, Nicholas J, Torous J, Birrell L, Li E, et al. Using science to sell apps: Evaluation of mental health app store quality claims. NPJ Digit Med 2019 Mar 22;2(1):18 [FREE Full text] [doi: 10.1038/s41746-019-0093-1] [Medline: 31304366]
29. Huguet A, Rao S, McGrath PJ, Wozney L, Wheaton M, Conrod J, et al. A Systematic Review of Cognitive Behavioral Therapy and Behavioral Activation Apps for Depression. PLoS One 2016;11(5):e0154248 [FREE Full text] [doi: 10.1371/journal.pone.0154248] [Medline: 27135410]
30. Buechi R, Faes L, Bachmann LM, Thiel MA, Bodmer NS, Schmid MK, et al. Evidence assessing the diagnostic performance of medical smartphone apps: a systematic review and exploratory meta-analysis. BMJ Open 2017 Dec 14;7(12):e018280 [FREE Full text] [doi: 10.1136/bmjopen-2017-018280] [Medline: 29247099]
31. Huckvale K, Venkatesh S, Christensen H. Toward clinical digital phenotyping: a timely opportunity to consider purpose, quality, and safety. NPJ Digit Med 2019 Sep 6;2(1):88 [FREE Full text] [doi: 10.1038/s41746-019-0166-1] [Medline: 31508498]
32. Huckvale K, Torous J, Larsen ME. Assessment of the Data Sharing and Privacy Practices of Smartphone Apps for Depression and Smoking Cessation. JAMA Netw Open 2019 Apr 05;2(4):e192542 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.2542] [Medline: 31002321]
33. Nicholas J, Shilton K, Schueller SM, Gray EL, Kwasny MJ, Mohr DC. The Role of Data Type and Recipient in Individuals' Perspectives on Sharing Passively Collected Smartphone Data for Mental Health: Cross-Sectional Questionnaire Study. JMIR Mhealth Uhealth 2019 Apr 05;7(4):e12578 [FREE Full text] [doi: 10.2196/12578] [Medline: 30950799]
34. Martinez-Martin N, Insel TR, Dagum P, Greely HT, Cho MK. Data mining for health: staking out the ethical territory of digital phenotyping. NPJ Digit Med 2018 Dec 19;1(1):1-5 [FREE Full text] [doi: 10.1038/s41746-018-0075-8] [Medline: 31211249]

35. Hahn T, Nierenberg AA, Whitfield-Gabrieli S. Predictive analytics in mental health: applications, guidelines, challenges and perspectives. Mol Psychiatry 2017 Jan 15;22(1):37-43. [doi: 10.1038/mp.2016.201] [Medline: 27843153]
36. Byambasuren O, Sanders S, Beller E, Glasziou P. Prescribable mHealth apps identified from an overview of systematic reviews. NPJ Digit Med 2018 May 9;1(1):12 [FREE Full text] [doi: 10.1038/s41746-018-0021-9] [Medline: 31304297]
37. US Geological Survey. 2019. Data Management-Data Standards URL: https://www.usgs.gov/products/data-and-tools/data-management/data-standards [accessed 2019-10-01]
38. Huckvale K, Prieto JT, Tilney M, Benghozi P, Car J. Unaddressed privacy risks in accredited health and wellness apps: a cross-sectional systematic assessment. BMC Med 2015 Sep 07;13:214 [FREE Full text] [doi: 10.1186/s12916-015-0444-y] [Medline: 26404673]
39. DeviceAtlas. 2019. Android v iOS market share 2019 URL: https://deviceatlas.com/blog/android-v-ios-market-share#uk [accessed 2019-09-20]
40. Junqué de Fortuny E, Martens D, Provost F. Predictive Modeling With Big Data: Is Bigger Really Better? Big Data 2013 Dec;1(4):215-226. [doi: 10.1089/big.2013.0037] [Medline: 27447254]
41. Barnett S, Vasa R, Tang A. A conceptual model for architecting mobile applications. 2015 Presented at: 12th Working IEEE/IFIP Conference on Software Architecture; 4-8 May; Montreal, Canada. [doi: 10.1109/wicsa.2015.28]
42. Barnett S. Extracting technical domain knowledge to improve software architecture. Melbourne, Australia: Swinburne University of Technology; 2018. URL: https://researchbank.swinburne.edu.au/items/458ce762-3ba9-441c-a389-933640d828fb/1/ [accessed 2019-10-25]
43. Shye A, Scholbrock B, Memik G. Into the wild: studying real user activity patterns to guide power optimizations for mobile architectures. 2009 Presented at: 42nd Annual IEEE/ACM International Symposium on Microarchitecture (MICRO); 12-16 Dec; New York, USA. [doi: 10.1145/1669112.1669135]
44. Vasa R. Semantic Scholar. Melbourne, Australia: Swinburne University of Technology; 2010. Growth and change dynamics in open source software systems URL: https://www.semanticscholar.org/paper/Growth-and-change-dynamics-in-open-source-software-Vasa/d3ff86d4ddeefd681a058c84d625f9215719442f [accessed 2019-10-25]
45. Fowler M, Beck K, Brant J, Opdyke W, Roberts D. Refactoring: Improving The Design Of Existing Code. Boston, Massachusetts, USA: Addison-Wesley Professional; 1999.
46. Bass L, Clements P, Kazman R. Software Architecture In Practice (3rd Edition). Boston, Massachussets, USA: Addison-Wesley Professional; 2019.
47. Christensen H. ANZCTR. 2019. The Future Proofing Study: A School-Based Depression Prevention Trial URL: https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=377664&isReview=true [accessed 2019-09-20]
48. Demchenko Y, de Laat C, Membrey P. Defining architecture components of the big data ecosystem. 2014 Presented at: 2014 International Conference on Collaboration Technologies and Systems (CTS); 19-23 May; Minneapolis, Minnesota, USA. [doi: 10.1109/cts.2014.6867550]
49. Dijkstra E. Selected Writings On Computing: A Personal Perspective. New York, USA: Springer; 2019.
50. Mollah M, Azad M, Vasilakos A. Security and privacy challenges in mobile cloud computing: Survey and way ahead. Journal of Network and Computer Applications 2017 Apr;84:38-54 [FREE Full text] [doi: 10.1016/j.jnca.2017.02.001]
51. Naab M, Braun S, Lenhart T, Hess S, Eitel A, Magin D. Why data needs more attention in architecture design - experiences from prototyping a large-scale mobile app ecosystem. 2015 Presented at: 12th Working IEEE/IFIP Conference on Software Architecture; 4-8 May; Montreal, Canada. [doi: 10.1109/wicsa.2015.13]
52. Martínez-Pérez B, de la Torre-Díez I, López-Coronado M. Privacy and security in mobile health apps: a review and recommendations. J Med Syst 2015 Jan;39(1):181. [doi: 10.1007/s10916-014-0181-3] [Medline: 25486895]
53. Chassang G. The impact of the EU general data protection regulation on scientific research. Ecancermedicalscience 2017;11:709 [FREE Full text] [doi: 10.3332/ecancer.2017.709] [Medline: 28144283]
54. Carter A, Liddle J, Hall W, Chenery H. Mobile Phones in Research and Treatment: Ethical Guidelines and Future Directions. JMIR Mhealth Uhealth 2015 Oct 16;3(4):e95 [FREE Full text] [doi: 10.2196/mhealth.4538] [Medline: 26474545]
55. Sila-Nowicka K, Thakuriah P. The trade-off between privacy and geographic data resolution. A case of GPS trajectories combined with the social survey results. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci 2016 Jun 08;XLI-B2:535-542. [doi: 10.5194/isprs-archives-xli-b2-535-2016]
56. https://awareframework.com/. Open-source context instrumentation framework for everyone URL: https://awareframework.com/ [accessed 2019-10-01]
57. Torous J, Kiang MV, Lorme J, Onnela J. New Tools for New Research in Psychiatry: A Scalable and Customizable Platform to Empower Data Driven Smartphone Research. JMIR Ment Health 2016 May 05;3(2):e16 [FREE Full text] [doi: 10.2196/mental.5165] [Medline: 27150677]
58. Onnela Lab. 2019. Beiwe Research Platform URL: https://www.beiwe.org/ [accessed 2019-10-01]
59. Lind MN, Byrne ML, Wicks G, Smidt AM, Allen NB. The Effortless Assessment of Risk States (EARS) Tool: An Interpersonal Approach to Mobile Sensing. JMIR Ment Health 2018 Aug 28;5(3):e10334 [FREE Full text] [doi: 10.2196/10334] [Medline: 30154072]

60. Saeb S, Zhang M, Karr CJ, Schueller SM, Corden ME, Kording KP, et al. Mobile Phone Sensor Correlates of Depressive Symptom Severity in Daily-Life Behavior: An Exploratory Study. J Med Internet Res 2015 Jul 15;17(7):e175 [FREE Full text] [doi: 10.2196/jmir.4273] [Medline: 26180009]
61. Faurholt-Jepsen M, Vinberg M, Frost M, Christensen EM, Bardram J, Kessing LV. Daily electronic monitoring of subjective and objective measures of illness activity in bipolar disorder using smartphones--the MONARCA II trial protocol: a randomized controlled single-blind parallel-group trial. BMC Psychiatry 2014 Nov 25;14:309 [FREE Full text] [doi: 10.1186/s12888-014-0309-5] [Medline: 25420431]
62. Audacious Software. 2018. Passive Data Kit URL: https://passivedatakit.org/ [accessed 2019-10-01]
63. thehyve. RADAR-base - an mHealth platform for generalised data collection URL: https://thehyve.nl/solutions/radar-base/ [accessed 2019-10-01]
64. Onnela Lab. 2019. Using Beiwe URL: https://www.beiwe.org/#panel3 [accessed 2019-10-01]
65. Onnela Lab. 2017. Beiwe data privacy and security URL: http://wiki.beiwe.org/wiki/Beiwe_Data_Privacy_and_Security [accessed 2019-10-01]
66. Jackson A, Cavanagh J, Scott J. A systematic review of manic and depressive prodromes. J Affect Disord 2003 May;74(3):209-217. [Medline: 12738039]
67. Addington J. The prodromal stage of psychotic illness: observation, detection or intervention? J Psychiatry Neurosci 2003 Mar;28(2):93-97 [FREE Full text] [Medline: 12670126]

Abbreviations

eHealth: electronic health
e-mental health: electronic mental health
InSTIL: Intelligent Sensing to Inform and Learn
IRB: Institutional Review Board
PHQ-9: Patient Health Questionnaire

Edited by G Eysenbach; submitted 25.09.19; peer-reviewed by D Mohr; comments to author 12.10.19; revised version received 22.10.19; accepted 22.10.19; published 06.11.19.

Please cite as:
Barnett S, Huckvale K, Christensen H, Venkatesh S, Mouzakis K, Vasa R
Intelligent Sensing to Inform and Learn (InSTIL): A Scalable and Governance-Aware Platform for Universal, Smartphone-Based Digital Phenotyping for Research and Clinical Applications
J Med Internet Res 2019;21(11):e16399
URL: https://www.jmir.org/2019/11/e16399
doi:10.2196/16399
PMID:31692450

Viewpoint

The Last Mile: Where Artificial Intelligence Meets Reality

Enrico Coiera[1], MBBS, PhD

Australian Institute of Health Inovation, Macquarie University, Sydney, Australia

Corresponding Author:
Enrico Coiera, MBBS, PhD
Australian Institute of Health Inovation
Macquarie University
Level 6 75 Talavera Rd
Sydney, 2109
Australia
Phone: 61 298502403
Email: enrico.coiera@mq.edu.au

Related Article:

Comment on: http://jmir.org/2019/10/e16194/

Abstract

Although much effort is focused on improving the technical performance of artificial intelligence, there are compelling reasons to focus more on the implementation of this technology class to solve real-world applications. In this "last mile" of implementation lie many complex challenges that may make technically high-performing systems perform poorly. Instead of viewing artificial intelligence development as a linear one of algorithm development through to eventual deployment, there are strong reasons to take a more agile approach, iteratively developing and testing artificial intelligence within the context in which it finally will be used.

(J Med Internet Res 2019;21(11):e16323) doi:10.2196/16323

KEYWORDS

artificial intelligence; implementation sceince; sociotechnical systems

Trying to separate information technology from the way it is used by people is like trying to separate a breath from the lungs that took it. We are inextricably bound to technology, and it shapes us as much as we shape it. This idea that humans and machines together constitute a larger sociotechnical system is foundational to our understanding of the modern world [1].

The sociotechnical lessons that the past hold for emerging technologies such as artificial intelligence (AI) should come as no surprise [2]. Amongst these old lessons is the maxim that the application of technology should be shaped by the problem at hand, and not the technology [3]. Although innovation in technology opens up new classes of solution, only the real world can tell us which problems are most worth solving and which solution class is most appropriate. The corollary is that a technology-driven demonstration of success with a problem does not necessarily mean it is the best solution for that problem, nor even that the problem is important.

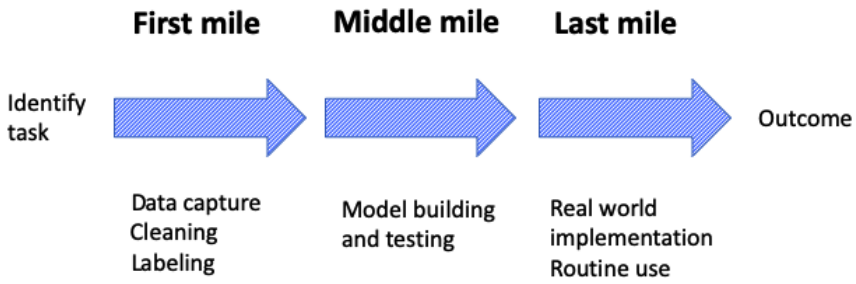
There are thus two broad agendas for AI research [4]. The first is a technical one, pursuing new methods, architectures, and technologies that push the boundary of machine intelligence. The second is an applied agenda, where we focus on questions of implementation, extending our understanding of how best to exploit technology within the sociotechnical world it must operate.

The risk we pay for technology neophilia, our love of the new, is that we may not see the system for the technology. With AI, much research focuses on demonstrating technological success on clinical tasks such as image interpretation, diagnosis, or treatment options [5]. There are many compelling reasons for us to shift the balance from such technical demonstrations to real-world applications, not least because many technical leaps are still needed to solve challenging real-world problems.

There are three broad stages in the development of data-driven technologies like machine learning (Figure 1). Once a task such as diagnosis has been selected for algorithm development, in the "first mile," data are acquired, possibly labelled, and preprocessed or "cleaned." Next is the middle mile, which focusses on developing and testing the technical performance of different algorithms built using these data. Only in the last

mile are algorithms embedded in real-world processes and tested for impact on real-world outcomes.

Figure 1. Three stages in the development of artificial intelligence technologies.



Each "mile" has its own challenges. "First mile problems" include the foundational challenges of gathering and curating high-quality data. For technologies such as machine learning, which are often dependent on large amounts of high-quality data, a bottleneck in data acquisition translates into a roadblock to technology application. The middle mile is home to all the challenges of data-driven algorithm development, including managing biases, replicability, causal inference, avoiding overfitting on training data, and enhancing the generalizability of any models and algorithms developed.

In the last mile, we face the reality that AI does not do anything on its own. It must be connecting somehow to real-world processes, and its impact on those processes needs to be consequential. It is at this point that technology developed for its own sake quickly comes to grief. For example, it is one thing to demonstrate machine learning can interpret thyroid scans for cancer as well or better than humans—a technical feat [6]; it is another for that feat to be meaningful. In the current setting, where thyroid cancer is both overdiagnosed and overtreated, we do not necessarily need better diagnoses. Instead, we need more nuanced and less aggressive approaches to management [7].

Last mile challenges are thus ones of implementation, and for researchers, of implementation science. These challenges exist at many levels, and include the following:

- *Measurement*: Standard metrics of AI performance relate to how well it completes its assigned task. We traditionally use performance measures like sensitivity, specificity, and area under the receiver operating characteristic curve (AUC). There is, however, a long chain of events that must occur between high technical performance and actual impact on clinical outcome. Success at an early stage in this information value chain is necessary, but not sufficient, to guarantee impact in the real world [8]. Evaluating real-world performance thus requires a shift from measuring technical accuracy to evaluating impact on processes and people. Measures like AUC can also be misleading in clinical settings [9]. AUC measures the overall performance of an algorithm across the entire receiver operating characteristic curve, while ideal real-world operation may best be limited to one segment of it [10].
- *Generalization and Calibration*: When an AI is trained on historic data, its future performance is dependent on how well new data match the historic data. It is a common implementation challenge to discover that a high-performing algorithm developed on data from one population deteriorates when used on a different population, reflecting underlying differences in the frequency and nature of events within the data sets. For this reason, AIs may need to be tuned to a specific end population. In many settings, this final population will also be dynamic, varying because of recurrent events like seasonal disease shifts, changes in population characteristics, and unexpected new events such as disease outbreaks. This means that AIs may need to be recalibrated, either periodically or dynamically, to reflect population changes. We also will need to closely monitor AI performance to detect shifts in its behavior that indicate that recalibration is needed [11]. More challenging is the fact that the more effectively an AI improves outcomes, the faster its performance may appear to degrade, as its very success can alter the association between predictors in its model and outcomes [12].
- *Local Context*: It is a fundamental tenet of implementation science that differences in the context in which a technology is embedded are associated with changes in performance. If we see an organization as a network of people, processes, and technologies, it is clear that the network underpinning any two organizations will be different. Implementation can be seen as the act of fitting a new technology or process into a pre-existing organizational network, and the goodness of fit of technology to network will shape any impact on organizational performance [13]. This is as true for AI as it is for digital health, in general, or for any new process or technology. To complicate matters, organizational networks are themselves dynamic. The impact of a technology will thus change with time, as the way it "fits" an organizational network changes and possibly because of its own presence—past connections will disappear or be replaced by new ones.

The software world has made a fundamental shift from seeing software development as a linear process that starts with user requirements and ends with end-user testing, to an agile one where users are embedded in a rapid and iterative process that adaptively fits software to users. Implementation science must travel the same path and especially so with dynamically changing AI.

AI development should not be seen as a linear journey stretching from the first to the last mile. Doing so risks the end product not meeting real-world needs, just as it does with software. Instead, implementation should be seen as an agile, iterative,

and lightweight process of obtaining training data, developing algorithms, and crafting these into tools and workflows. Finding the right balance between reusing general technology and building to meet local needs will be crucial [14]. Either way, AI should not be created far from the place they will be used. Ideally, they should be born deep within the network that they will ultimately live in.

Conflicts of Interest

None declared.

References

1. Berg M, Aarts J, van der Lei J. ICT in health care: sociotechnical approaches. Methods Inf Med 2003;42(4):297-301. [Medline: 14534625]
2. Stefanelli M. The socio-organizational age of artificial intelligence in medicine. Artif Intell Med 2001 Aug;23(1):25-47. [Medline: 11470215]
3. Coiera E. Medical informatics. BMJ 1995 May 27;310(6991):1381-1387. [doi: 10.1136/bmj.310.6991.1381] [Medline: 7787545]
4. Coiera EW. Artificial intelligence in medicine: the challenges ahead. J Am Med Inform Assoc 1996;3(6):363-366 [FREE Full text] [doi: 10.1136/jamia.1996.97084510] [Medline: 8930853]
5. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med 2019 Jan;25(1):44-56. [doi: 10.1038/s41591-018-0300-7] [Medline: 30617339]
6. Li X, Zhang S, Zhang Q, Wei X, Pan Y, Zhao J, et al. Diagnosis of thyroid cancer using deep convolutional neural network models applied to sonographic images: a retrospective, multicohort, diagnostic study. The Lancet Oncology 2019 Feb;20(2):193-201. [doi: 10.1016/S1470-2045(18)30762-9]
7. Coiera E. On algorithms, machines, and medicine. Lancet Oncol 2019 Feb;20(2):166-167. [doi: 10.1016/S1470-2045(18)30835-0] [Medline: 30583847]
8. Coiera E. Assessing Technology Success and Failure Using Information Value Chain Theory. Stud Health Technol Inform 2019 Jul 30;263:35-48. [doi: 10.3233/SHTI190109] [Medline: 31411151]
9. Cook NR. Use and Misuse of the Receiver Operating Characteristic Curve in Risk Prediction. Circulation 2007 Feb 20;115(7):928-935. [doi: 10.1161/CIRCULATIONAHA.106.672402]
10. Thomas G, Kenny LC, Baker PN, Tuytten R. A novel method for interrogating receiver operating characteristic curves for assessing prognostic tests. Diagn Progn Res 2017 Nov 15;1(1):17. [doi: 10.1186/s41512-017-0017-y]
11. Ong M, Magrabi F, Coiera E. Syndromic surveillance for health information system failures: a feasibility study. J Am Med Inform Assoc 2013 May 01;20(3):506-512 [FREE Full text] [doi: 10.1136/amiajnl-2012-001144] [Medline: 23184193]
12. Lenert M, Matheny ME, Walsh CG. Prognostic models will be victims of their own success, unless…. J Am Med Inform Assoc 2019 Sep 03:ocz145. [doi: 10.1093/jamia/ocz145] [Medline: 31504588]
13. Coiera E. Chapter 12: Implementation, in Guide to Health Informatics. Boca Raton: CRC Press; 2016:173-194.
14. Johannessen L, Ellingsen G. Integration and Generification—Agile Software Development in the Healthcare Market. Comput Supported Coop Work 2009 Sep 16;18(5-6):607-634. [doi: 10.1007/s10606-009-9097-8]

Abbreviations

AI: artificial intelligence
AUC: area under the receiver operating characteristic curve

Edited by G Eysenbach; submitted 18.09.19; peer-reviewed by D Mendes; accepted 20.10.19; published 08.11.19.

Please cite as:
Coiera E
The Last Mile: Where Artificial Intelligence Meets Reality
J Med Internet Res 2019;21(11):e16323
URL: http://www.jmir.org/2019/11/e16323/
doi:10.2196/16323
PMID:31702559

Viewpoint

Unlocking the Power of Artificial Intelligence and Big Data in Medicine

Christian Lovis[1,2], MD, MPH, FACMI

[1]Division of Medical Information Sciences, University Hospitals of Geneva, Geneva, Switzerland

[2]Department of Radiology and Medical Informatics, University of Geneva, Geneva, Switzerland

Corresponding Author:
Christian Lovis, MD, MPH, FACMI
Division of Medical Information Sciences
University Hospitals of Geneva
Gabrielle Perret Gentil 4
Geneva, 1205
Switzerland
Phone: 41 22 37 26201
Email: Christian.Lovis@hcuge.ch

Abstract

Data-driven science and its corollaries in machine learning and the wider field of artificial intelligence have the potential to drive important changes in medicine. However, medicine is not a science like any other: It is deeply and tightly bound with a large and wide network of legal, ethical, regulatory, economical, and societal dependencies. As a consequence, the scientific and technological progresses in handling information and its further processing and cross-linking for decision support and predictive systems must be accompanied by parallel changes in the global environment, with numerous stakeholders, including citizen and society. What can be seen at the first glance as a barrier and a mechanism slowing down the progression of data science must, however, be considered an important asset. Only global adoption can transform the potential of big data and artificial intelligence into an effective breakthroughs in handling health and medicine. This requires science and society, scientists and citizens, to progress together.

(J Med Internet Res 2019;21(11):e16607) doi:10.2196/16607

KEYWORDS

medical informatics; artificial intelligence; big data

Introduction

Most of the daily news and recently published scientific papers on research, innovations, and applications in artificial intelligence (AI) refer to what is known as machine learning—algorithms using massive amounts of data and various methodologies to find patterns, support decisions, make predictions, or, for the deep learning part, self-identify important features in data. However, AI is a complex concept to grasp, and most people have little understanding of what it really is. AI was founded as an academic discipline in 1956 and, despite its youth, already has a rich history [1,2]. In more than 60 years of exploration and progress, AI has become a large field of research and development involving multidisciplinary approaches to address many challenges, from theoretical frameworks, methods, and tools to real implementations, risk analysis, and impact measures. The definition of AI is a moving target and changes over time with the evolution of the field. Since its early days, the field of AI has allowed the development of many techniques supporting decision support and prediction, as it is usually made by humans. As early as 1958, a perceptron was expected to be able "to walk, talk, see, write, reproduce itself and be conscious of its existence," which led a large scientific controversy between neural network and symbolic reasoning approaches [3]. The landscape of AI research includes knowledge representation and engineering; rule-based and symbolic reasoning; temporal reasoning and planning; sensing and perception; learning; evolutionary, emerging, social behaviors; and the ability to move and manipulate objects, to name the most important [4]—deep machine learning with autonomous features extraction. It is the point of view taken in the paper, acknowledging however that, more recently, there is a trend to restrict AI to the latter, autonomous deep machine learning. As a consequence of the wide landscape, the field draws at large through philosophy, mathematics, information sciences, computer science, psychology, anthropology, social sciences, linguistics, and many others. For some experts and

visionary people such as Ray Kurzweil, deep machine learning will allow building of an artificial general intelligence that is able to develop itself autonomously and to have the capacity to understand or learn any intellectual task that a human being can, and even go far beyond the limits of human intelligence [5], but most experts would agree that there are some big missing pieces and it is still a long way off, despite recent potential important advances in quantic computing [6]. A recent white paper published by the European Commission and authored by the members of the High-Level Expert Group on AI provides, in a few pages, a good overview on what AI is, its main capabilities, applicable expectations, and disciplines involved [7].

Taking into the field of AI at large, it is important to emphasize that AI is already broadly used today in medicine. Decision support based on knowledge engineering and rule-based systems are implemented widely in computerized provider order entry (CPOE) worldwide. Advanced signal processing is implemented in pacemakers or defibrillators to take decisions, in cochlear-implants with man-machine interfaces, in electrocardiograms to provide signal analysis and automated diagnosis, etc.

The AI field in itself is aspirational and is expected to contribute significantly to medicine, from research to citizen-centered health. Machine learning and deep learning has led most recent major breakthroughs in AI, such as sound (speech and music) recognition and image (face, radiology, pathology, dermatology, etc) recognition, and in gaming. Recently, image recognition has almost reached a level of maturity through which it can be used and developed by nonexperts in AI [8,9]. However, the hype around AI in these last few years has built high expectations and similarly high fears. There are still very few systems based on autonomous deep learning that have emerged widely in the commercial market.

The world of AI could roughly be summarized in three sequential and superposed acts:

Act 1: Humans teach machines to handle data and information.

Act 2: Humans teach expertise to machines.

Act 3: Humans teach machines to learn alone.

Challenges

There are many challenges that need to be addressed in the field of AI, when it comes to medicine. Most of them are not exclusive to medicine and health, but their addition makes the goals significantly much harder to reach.

Bayesian Trap

Medicine and health determinants, in general, are characterized by their usually fundamental Bayesian property. In the Bayesian probability approach, a prior probability is required to evaluate the strength of the prediction.

Most of what is used in medicine, notably but not exclusively, to establish diagnosis, falls in the Bayesian approach. For example, if a person has a fever, what is the probability of it being the flu? If a person has a high measurement of blood sugar, what is the probability of it being diabetes? To illustrate the Bayesian trap, let us take a simple example—a pregnancy test. This is a simple test; it can be positive or negative. Let us imagine that we use a classical test, which has 99% sensitivity and 95% specificity. If 100 tests are performed for 100 persons and 5 turn out to be positive, the question is, know how many of them turn out to be pregnant women. This question is defined as determining the positive predictive value of a positive test, and it gives the probability of a positive test to be really signing the presence of the factor the test is testing. To answer this question, we need to know the prior probability of being pregnant in the tested population. To understand this, imagine that 100% men were tested; in this case, none of the 5 positive tests would correspond to a pregnant woman. Similarly, if all persons tested are pregnant woman, then all 5 positive tests would correspond to pregnant women. If the prior probability is around 1%, then applying the Bayesian rules returns that the probability to be pregnant when a test is reported positive is about 17%. This means that about 4 of 5 tests are false positive. At the other end, if the prior probability is around 20% (ie, a woman with several factors suggesting a potential pregnancy), the probability of a positive test to be a true positive is above 80%. Thus, less than one test out of five is a false-positive. The example shows the major consequences of the prior probability in Bayesian situations.

These are the consequences of AI. The models must take into account prior probability in the population they are used. This should be better understood even when reporting results in the literature, often limited to specificity and sensitivity. Another consequence only becomes visible when several focused and near-to-perfect systems are used together in complex cases. For example, having many systems, each with its own false-positive rate can end up with consolidated systems that have the sum of all false-positives. This has been shown well with decision support system in CPOE, with a very high rate of false-positive alerts, especially with patients receiving complex drug therapies [10,11].

Regulatory Labyrinth

Most diagnostic or therapeutic means used nowadays in medicine have to go through complex regulatory frameworks to get market approval. The regulatory agencies mostly base their decisions on safety, evidence, and added value. In addition, medicoeconomic assessments are often used by health agencies according to various dimensions such as quality-adjusted life year and burden of the disease, by using indicators such as disability-adjusted life-years [12-14]. These decisions thus have economical and legal consequences, including accountability. The role of regulatory agencies is discussed, especially around topics that are getting into the market, such as in image recognition [15]. For example, a call for inputs for “Artificial Intelligence and Machine Learning in Software as a Medical Device“ has been launched by the Federal Drug Administration [16]. This is an important aspect, as regulatory agency support is an important asset in building trust for most care professionals to use medical tools and for companies to invest in robust products ready for the market. However, this requires us to define a clear regulatory framework, appropriate evaluation

processes, and benchmark tools without blocking innovation [17].

Education and Practice Gap

Medicine is a science with numerous tools and devices, from stethoscopes to scalpels, microscope to scanners, scores, guidelines, etc. Most of these tools and devices require education and sometimes very specific certification processes for care professionals that use them, not to speak about a good experience. This should be also the case for software, algorithms, and other decision support systems. However, this is not the case. Education to use software and understand systems as important as the computerized patient records is often minimal. When it is about big data and AI, education on the topic is worse, usually inexistent. There are only very few medical schools that teach the use of AI to future health professionals. AI should become mandatory teaching in all medical schools in the world as a priority. Experts have been raising the question since 20 years, but it has received real focus only recently [18-22]. In 5-10 years, when current young students will be starting their clinical activities, machine learning based on data science will have become embedded in many activities, devices, and software and its use, misuse, and overuse and consequences on patients and accountability will depend on how users will master it [23].

Data Quality Chiasm

Data quality is a recurring topic of discussion when it comes to big data and analytics. One of the characteristics of the big data era is that data are often used for a purpose that differs from the one that motivated data acquisition. This is a notable difference with traditional hypothetic-deductive scientific approaches in medicine, where a hypothesis leads to a methodology design, which itself will lead to specific data acquisition. In the big data era, the primary goal of data-producing processes is often completely independent from possible use of the data. It is interesting to emphasize that long-term clinical cohorts and long-term biobanks face similar challenges. Designing long-term cohorts and building metadata framework and standard operating procedures for biobanking are important challenges, as they have to project usages that will be made years after the initial design.

These questions have led to a consequent literature addressing the question of data quality and secondary usage of clinical data. However, most of this work tries to describe dimensions able to assess the "intrinsic" or "absolute" quality of data [24-28]. Another approach could be to adopt a "fit-for-purpose" approach, which considers only the quantitative and descriptive properties of data, allowing further processing. The "qualitative" properties of any dataset can only be assessed in conjunction with a specific secondary usage. This means that the same dataset will be appropriate to answer some scientific questions and not appropriate to answer others. The data are not "good" or "bad" by themselves; they are "good" or "bad" when used in a specific context: the "fit-for-purpose" assessment. This is one of the major objectives of the FAIR data initiative, which aims at insuring "*a posteriori*" data usability (see below).

An unexpected consequence of the "data quality chiasm" is its influence on modifying acquisition processes, especially in clinical contexts. One often hears sentences such as "the quality of clinical data is not good enough for research." As a result, there is a constant pressure to move toward more structured data acquisition processes. For example, the RECIST (Response Evaluation Criteria In Solid Tumors) guidelines are meant to standardize the radiologic evaluation criteria in solid tumors oncological trial treatments. This has been successfully developed for trials. Use of RECIST requires good experience to avoid interobserver variability, which can be as high as 20% [29-31]. This assessment has been adapted to reflect changes in radiological response, for example, in immunotherapies where the size of tumors can increase despite good therapeutic response [32]. Unfortunately, there is growing pressure to extend the use of RECIST and other similarly structured staging guidelines beyond clinical trials for all radiological staging to improve the capacity to use standard clinical care for therapeutic assessment. As a consequence, this leads to a high time pressure on operational activities of radiology departments and an increasing number of inexperienced people using these types of staging. With the progression of natural interfaces such as voice recognition and natural language processing and their increased daily use in a growing number devices, I would argue in favor of avoiding artificial structuring many data acquisition processes and keep the data in their most natural form, exploiting more natural interactions such as voice and text and developing strong natural language processing tools that can be applied to produce structured information in a postprocessing step. This will allow reprocessing of all narratives whenever needed by new structured resources required.

Quest for Truth

Many aspects of the landscape of artificial intelligence require a good idea of what is true. Knowledge engineering builds the graph of the "known" or the "relevant" such as it is made in SNOMED CT (Systematized Nomenclature of Medicine - Clinical Terms) or the Open Biological and Biomedical Ontology Foundry [33,34]. The same applies with rule-based techniques or symbolic reasoning, which need to be able to express rules, that is, truth in a formalized way, but also in supervised machine learning approaches, which require having training sets that express truth, at least a probabilistic truth. There are a lot of expectations in these approaches, especially when combing them [35,36], but all of them, except unsupervised deep machine learning, require some sources of truth, which leads to the fundamental question of finding the sources of truth in life sciences and the level of evidence supporting that truth. At the first glance, it seems to be a trivial question. However, the "truth" is often "lost in text" because for most of it, the sources rely on complex narratives that contextualize the messages they convey. In addition, the "truth" is very diluted. For example, with more than 2500 papers indexed daily in Medline/PubMed [37], it is nearly impossible for an expert to catch everything published in its own research field. Finally, and by nature, science is evolving, and thus, scientific "truth" of what was true once may no more be true today. For example, it was clear until recently that there are two

types of lymphocytes—the B and T cells. However, a recent paper from Rizwan et al [38] describes a new type of lymphocyte, bearing characteristics of both B and T cells, which may play a role in driving autoimmunity in some diseases such as diabetes [38]. Sources of truth and their characterizations, such as the level of evidence or their context of use, are increasingly important. This should be available to all, similar to Cochrane [39], covering all area of life sciences; maintained; and in machine-readable form.

Building Trust

In science, trust is strongly related to building evidence. Trust is important in not only the scientific community, but also at large, to build adoption, political support, and public acceptance. In summer 2019, a survey published by the Pew Research Center showed a positive trend among the public: science acts for the good, but with concerns about integrity, transparency, and bias. Overall, 86% of Americans say they have at least "a fair amount" of confidence in scientists [40]. One of the challenges is that scientific reliability has often been confused with trustworthiness [41]. Scientific evidence can be very strong, such as for immunization or Web-based health information, but the trust can be much lower [42,43]. There are many dimensions that have been discussed in building trust in science, but they can be summarized in three concepts, one for the scientists and the organizations, one for the objects of the research, and one for the processes. Integrity is first and most important and covers scientific integrity, funding, conflict of interests, etc. Transparency must be present for the motivation, outcomes, and process. Finally, methodologies applied to handle the processes must be strong and robust. Building evidence requires many dimensions to be taken into account, such as bias, generalizability, reproducibility, and explainability. Some challenges are more difficult in big data and AI. Proper control of data acquisition and flow is usually more difficult than that in traditional controlled studies. The consequence is that the data have specific properties, which are not always well managed, such as selection biases. Sometimes, the assumptions constraining the use of analytical tools are not well understood, such as homoscedasticity for many statistical tests. In addition, deep machine learning is facing the challenges of precise reproducibility and explainability. The latter is currently the object of numerous works, trying to understand intermediate representation of data in neural networks that can predict and explain their behavior. Explainability and interpretability are often used interchangeably. Interpretability is the extent to which it is possible to predict how the system will behave, given a change in input or algorithmic parameters. On the other hand, explainability is the extent to which the internal mechanics of the deep learning system can be understood and thus explained. Molnar [44] published a very good overview of the problem in an open book available on GitHub. However, explainability might not be the best road to raise global trust in deep machine learning approaches, especially when the explanations themselves are hard to explain. Some other dimensions such as transparency, reproducibility, or uncertainty qualifications might be more effective [45]. For example, in *Science* in 2018, Hutson [46] reported a survey of 400 artificial intelligence papers presented at major conferences, with only 6% including code for the algorithms and 30% test data, thus considerably limiting reproducibility possibilities [46].

FAIR Data Hope

The FAIR Guiding Principles are guidelines to make data discoverable and processable by both humans and machines. They were first published by Wilkinson et al [47]. The FAIR Guiding Principles are based on a set of criteria listed in Textbox 1:

Textbox 1. FAIR data criteria.

Findable

- (Meta)data are assigned a globally unique and persistent identifier
- Data are described with rich metadata (defined below)
- Metadata clearly and explicitly include the identifier of the data described
- (Meta)data are registered or indexed in a searchable resource

Accessible

- (Meta)data are retrievable by their identifier using a standardized communications protocol
- The protocol is open, free, and universally implementable
- The protocol allows for an authentication and authorization procedure, where necessary
- Metadata are accessible, even when the data are no longer available

Interoperable

- (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation
- (Meta)data use vocabularies that follow FAIR Principles
- (Meta)data include qualified references to other (meta)data

Reusable

Meta(data) are richly described with a plurality of accurate and relevant attributes:

- (Meta)data are released with a clear and accessible data usage license
- (Meta)data are associated with detailed provenance
- (Meta)data meet domain-relevant community standards

Several frameworks have been defined to assess and evaluate the compliance to FAIR criteria, such as the FAIR maturity tools [48,49]. As such, FAIR data do not imply that data are in the Open Data space [50]. Access can be restricted, such as in the Harvard Dataverse, and this is an important point to emphasize. There might be a lot of restriction to have data or metadata available in the Open Data space, because of national regulation, privacy protection, intellectual property, etc. FAIR data do not make data available; they make data usable under the condition that it is authorized.

The FAIR initiative is crucial. It illustrates the movement of data from objects to assets initiated this last decade and described in the essay of Sabina Leonelli [51] recently published in *Nature*.

The initiative promotes the use of rich metadata framework, compliant to standards and formal descriptions. It promotes the use of free and open resources for descriptive information and protocols. It allows us to build a framework of shareable resources that can be used for processing, without actually sharing the data in an open space. FAIR allows building of a framework that is inclusive with all data sources, including those that are subject to authorization, with clear and open protocols.

Privacy – New Deal

In the era of big data, privacy requires special attention. Usual paradigms of limiting access to deidentified information are becoming less effective to protect privacy. Increasing heterogeneous data sources and richness of data about each of us, associated with data linkage techniques, strongly increases the possibility of reidentification, including anonymized data [52-57]. The challenge and potential impacts are even bigger for genetic information [58-60]. There is no good technical solution that can harmonize the challenge of preserving privacy and answering the increasing need of data-driven science for accessing large genomic et phenotypic datasets, and there are many ongoing ethical and legal discussions [61-66]. Interestingly, this is not restricted to science, and the same applies to patients' needs for health information [67]. There is a need for better global education about implications and risks of privacy, citizen, policy makers, students, research community, and all stakeholders. A recent scoping review [68] has shown that the understanding of anonymization and de-deidentification is heterogeneous in the scientific community [68]. Discrimination is one of the major risks in privacy breaches, and disclosing privacy information can have many consequences [69-71], including in reimbursement and insurance coverage [72,73]. It is important to find the right path between naïve positivism and irrational paranoia. An important step forward is to improve awareness and education of all stakeholders about privacy, technical limitations to protect it, and building regulatory barriers to avoid discrimination.

Conclusions

AI and big data in medicine are only in their childhood stages; they grow up fast. Whether they grow up well is still an open question that the future will answer. However, they will not grow up well without actively helping them do so. There are several important initiatives that contribute to this, such as the Global Alliance for Genomics and Health (GA4GH), an organization setting a policy and technical framework for respecting human rights to enable responsible genomic data sharing [74], or the European Union General Data Protection Regulation (GDPR) [75] that sets a completely novel privacy regulation for the European Union. Such initiatives are converging toward building a landscape that enables science while building trust in improving protection of individual rights. I invite the readers to visit the JMIR Open Access collections available on the Web on the following topics: “Big Data,” “Decision Support for Health Professionals,” and “Artificial Intelligence” [76-78].

Conflicts of Interest

None declared.

References

1. Bringsjord S, Govindarajulu N. Stanford Encyclopedia of Philosophy Archive. Stanford, CA: Metaphysics Research Lab, Stanford University; 2018. Artificial Intelligence URL: https://plato.stanford.edu/archives/fall2018/entries/artificial-intelligence/ [accessed 2019-09-22]
2. Wikipedia. 2019. Artificial intelligence URL: https://en.wikipedia.org/w/index.php?title=Artificial_intelligence&oldid=916837449 [accessed 2019-09-22]
3. Olazaran M. A Sociological Study of the Official History of the Perceptrons Controversy. Soc Stud Sci 2016 Jun 29;26(3):611-659. [doi: 10.1177/030631296026003005]
4. Luger G. Artificial intelligence: structures and strategies for complex problem solving. Boston, MA: Pearson Addison-Wesley; 2009:978.
5. Kurzweil R. The singularity is near: when humans transcend biology. New York: Viking; 2005:978.
6. Giles M. MIT Technology Review. Google researchers have reportedly achieved "quantum supremacy" URL: https://www.technologyreview.com/f/614416/google-researchers-have-reportedly-achieved-quantum-supremacy/ [accessed 2019-09-22]
7. High-Level EGOAI(. European Commission. A definition of Artificial Intelligence: main capabilities and scientific disciplines URL: https://ec.europa.eu/digital-single-market/en/news/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines [accessed 2019-09-22]
8. Godec P, Pancur M, Ilenic N, Copar A, Stražar M, Erjavec A, et al. Democratized image analytics by visual programming through integration of deep models and small-scale machine learning. Nat Commun 2019 Oct 7;10(1):7. [doi: 10.1038/s41467-019-12397-x]
9. Faes L, Wagner SK, Fu DJ, Liu X, Korot E, Ledsam JR, et al. Automated deep learning design for medical image classification by health-care professionals with no coding experience: a feasibility study. The Lancet Digital Health 2019 Sep;1(5):e232-e242. [doi: 10.1016/s2589-7500(19)30108-6]
10. Carli D, Fahrni G, Bonnabry P, Lovis C. Quality of Decision Support in Computerized Provider Order Entry: Systematic Literature Review. JMIR Med Inform 2018 Jan 24;6(1):e3. [doi: 10.2196/medinform.7170]
11. Jung M, Hoerbst A, Hackl WO, Kirrane F, Borbolla D, Jaspers MW, et al. Attitude of Physicians Towards Automatic Alerting in Computerized Physician Order Entry Systems. Methods Inf Med 2018 Jan 20;52(02):99-108. [doi: 10.3414/me12-02-0007]
12. Neumann PJ, Sanders GD. Cost-Effectiveness Analysis 2.0. N Engl J Med 2017 Jan 19;376(3):203-205. [doi: 10.1056/NEJMp1612619] [Medline: 28099837]
13. Sanders GD, Neumann PJ, Basu A, Brock DW, Feeny D, Krahn M, et al. Recommendations for Conduct, Methodological Practices, and Reporting of Cost-effectiveness Analyses. JAMA 2016 Sep 13;316(10):1093. [doi: 10.1001/jama.2016.12195]
14. Mokdad AH, Ballestros K, Echko M, Glenn S, Olsen HE, Mullany E, et al. The State of US Health, 1990-2016. JAMA 2018 Apr 10;319(14):1444. [doi: 10.1001/jama.2018.0158]
15. Allen B. The Role of the FDA in Ensuring the Safety and Efficacy of Artificial Intelligence Software and Devices. J Am Coll Radiol 2019 Feb;16(2):208-210. [doi: 10.1016/j.jacr.2018.09.007] [Medline: 30389329]
16. U.S. Food and Drug Administration. 2019 Jun 26. Artificial Intelligence and Machine Learning in Software as a Medical Device URL: http://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device [accessed 2019-09-22]
17. Parikh RB, Obermeyer Z, Navathe AS. Regulation of predictive analytics in medicine. Science 2019 Feb 22;363(6429):810-812 [FREE Full text] [doi: 10.1126/science.aaw0029] [Medline: 30792287]
18. Chan KS, Zary N. Applications and Challenges of Implementing Artificial Intelligence in Medical Education: Integrative Review. JMIR Med Educ 2019 Jun 15;5(1):e13930 [FREE Full text] [doi: 10.2196/13930] [Medline: 31199295]

19. Masters K. Artificial intelligence in medical education. Med Teach 2019 Sep;41(9):976-980. [doi: 10.1080/0142159X.2019.1595557] [Medline: 31007106]
20. Wartman SA, Combs CD. Reimagining Medical Education in the Age of AI. AMA J Ethics 2019 Feb 01;21(2):E146-E152 [FREE Full text] [doi: 10.1001/amajethics.2019.146] [Medline: 30794124]
21. Data Science Institute. Training Medical Students and Residents for the AI Future URL: https://www.acrdsi.org/Blog/Medical-schools-must-prepare-trainees [accessed 2019-09-22]
22. Lillehaug SI, Lajoie SP. AI in medical education--another grand challenge for medical informatics. Artif Intell Med 1998 Mar;12(3):197-225. [Medline: 9626957]
23. Kolachalama VB, Garg PS. Machine learning and medical education. NPJ Digit Med 2018 Sep 27;1(1):54 [FREE Full text] [doi: 10.1038/s41746-018-0061-1] [Medline: 31304333]
24. Stausberg J, Bauer U, Nasseh D, Pritzkuleit R, Schmidt C, Schrader T, et al. Indicators of data quality: review and requirements from the perspective of networked medical research. GMS Med Inform Biom Epidemiol 2019;15(1):05.
25. Weiskopf NG, Weng C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. J Am Med Inform Assoc 2013 Jan 01;20(1):144-151 [FREE Full text] [doi: 10.1136/amiajnl-2011-000681] [Medline: 22733976]
26. Weiner MG. Toward Reuse of Clinical Data for Research and Quality Improvement: The End of the Beginning? Ann Intern Med 2009 Sep 01;151(5):359. [doi: 10.7326/0003-4819-151-5-200909010-00141]
27. Sandhu E, Weinstein S, McKethan A, Jain SH. Secondary Uses of Electronic Health Record Data: Benefits and Barriers. The Joint Commission Journal on Quality and Patient Safety 2012 Jan;38(1):34-40. [doi: 10.1016/s1553-7250(12)38005-7]
28. Hersh W. Adding value to the electronic health record through secondary use of data for quality assurance, research, and surveillance. Am J Manag Care 2007 Jun;13(6 Part 1):277-278 [FREE Full text] [Medline: 17567224]
29. Jiang B, Zhou D, Sun Y, Wang J. Systematic analysis of measurement variability in lung cancer with multidetector computed tomography. Ann Thorac Med 2017;12(2):95-100 [FREE Full text] [doi: 10.4103/1817-1737.203750] [Medline: 28469719]
30. Bellomi M, De Piano F, Ancona E, Lodigiani AF, Curigliano G, Raimondi S, et al. Evaluation of inter-observer variability according to RECIST 1.1 and its influence on response classification in CT measurement of liver metastases. Eur J Radiol 2017 Oct;95:96-101. [doi: 10.1016/j.ejrad.2017.08.001] [Medline: 28987705]
31. Yoon SH, Kim KW, Goo JM, Kim D, Hahn S. Observer variability in RECIST-based tumour burden measurements: a meta-analysis. Eur J Cancer 2016 Jan;53:5-15. [doi: 10.1016/j.ejca.2015.10.014] [Medline: 26687017]
32. Seymour L, Bogaerts J, Perrone A, Ford R, Schwartz LH, Mandrekar S, RECIST working group. iRECIST: guidelines for response criteria for use in trials testing immunotherapeutics. Lancet Oncol 2017 Mar;18(3):e143-e152 [FREE Full text] [doi: 10.1016/S1470-2045(17)30074-8] [Medline: 28271869]
33. The OBO Foundry. URL: http://www.obofoundry.org/ [accessed 2019-10-05]
34. SNOMED International. URL: https://www.snomed.org/snomed-international/who-we-are [accessed 2018-06-27]
35. Skevofilakas M, Zarkogianni K, Karamanos B, Nikita K. A hybrid Decision Support System for the risk assessment of retinopathy development as a long term complication of Type 1 Diabetes Mellitus. Conf Proc IEEE Eng Med Biol Soc 2010;2010:6713-6716. [doi: 10.1109/IEMBS.2010.5626245] [Medline: 21096083]
36. Hoehndorf R, Queralt-Rosinach N. Data Science and symbolic AI: Synergies, challenges and opportunities. DS 2017 Oct 17:1-12. [doi: 10.3233/ds-170004]
37. NIH: U.S. National Library of Medicine. Detailed Indexing Statistics: 1965-2017 URL: https://www.nlm.nih.gov/bsd/index_stats_comp.html [accessed 2017-10-08]
38. Ahmed R, Omidian Z, Giwa A, Cornwell B, Majety N, Bell DR, et al. A Public BCR Present in a Unique Dual-Receptor-Expressing Lymphocyte from Type 1 Diabetes Patients Encodes a Potent T Cell Autoantigen. Cell 2019 May;177(6):1583-1599.e16. [doi: 10.1016/j.cell.2019.05.007]
39. Cochrane. URL: https://www.cochrane.org/ [accessed 2019-10-05]
40. Pew Research Center: Science & Society. Trust and Mistrust in Americans' Views of Scientific Experts URL: https://www.pewresearch.org/science/2019/08/02/trust-and-mistrust-in-americans-views-of-scientific-experts/ [accessed 2019-10-05]
41. Kerasidou A. Trust me, I'm a researcher!: The role of trust in biomedical research. Med Health Care Philos 2017 Mar;20(1):43-50 [FREE Full text] [doi: 10.1007/s11019-016-9721-6] [Medline: 27638832]
42. Larson HJ, Clarke RM, Jarrett C, Eckersberger E, Levine Z, Schulz WS, et al. Measuring trust in vaccination: A systematic review. Hum Vaccin Immunother 2018 Jul 03;14(7):1599-1609 [FREE Full text] [doi: 10.1080/21645515.2018.1459252] [Medline: 29617183]
43. Sbaffi L, Rowley J. Trust and Credibility in Web-Based Health Information: A Review and Agenda for Future Research. J Med Internet Res 2017 Jun 19;19(6):e218 [FREE Full text] [doi: 10.2196/jmir.7579] [Medline: 28630033]
44. Interpretable Machine Learning. 2019. URL: https://christophm.github.io/interpretable-ml-book/ [accessed 2019-10-18]
45. Begoli E, Bhattacharya T, Kusnezov D. The need for uncertainty quantification in machine-assisted medical decision making. Nat Mach Intell 2019 Jan 7;1(1):20-23. [doi: 10.1038/s42256-018-0004-1]
46. Hutson M. Artificial intelligence faces reproducibility crisis. Science 2018 Feb 16;359(6377):725-726. [doi: 10.1126/science.359.6377.725] [Medline: 29449469]

47. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 2016 Mar 15;3(1):160018 [FREE Full text] [doi: 10.1038/sdata.2016.18] [Medline: 26978244]
48. GitHub. FAIR Metrics 2019 URL: https://github.com/FAIRMetrics/Metrics [accessed 2019-10-05]
49. Union POTE. Publications Office of the European Union. Turning FAIR into reality: final report and action plan from the European Commission expert group on FAIR data URL: https://publications.europa.eu/en/publication-detail/-/publication/7769a148-f1f6-11e8-9982-01aa75ed71a1 [accessed 2019-07-04]
50. Harvard University. 2019. Towards an integrated and FAIR Research Data Management @Harvard Internet URL: https://scholar.harvard.edu/mercecrosas/presentations/towards-integrated-and-fair-research-data-management-harvard [accessed 2019-04-10]
51. Leonelli S. Data — from objects to assets. Nature 2019 Oct 15;574(7778):317-320. [doi: 10.1038/d41586-019-03062-w]
52. Malin B. Re-identification of familial database records. AMIA Annu Symp Proc 2006:524-528 [FREE Full text] [Medline: 17238396]
53. El Emam K, Jonker E, Arbuckle L, Malin B. A systematic review of re-identification attacks on health data. PLoS One 2011 Dec 2;6(12):e28071 [FREE Full text] [doi: 10.1371/journal.pone.0028071] [Medline: 22164229]
54. Dankar FK, El Emam K, Neisa A, Roffey T. Estimating the re-identification risk of clinical data sets. BMC Med Inform Decis Mak 2012 Jul 09;12(1):66 [FREE Full text] [doi: 10.1186/1472-6947-12-66] [Medline: 22776564]
55. Benitez K, Malin B. Evaluating re-identification risks with respect to the HIPAA privacy rule. J Am Med Inform Assoc 2010;17(2):169-177 [FREE Full text] [doi: 10.1136/jamia.2009.000026] [Medline: 20190059]
56. Barth-Jones D. The 'Re-identification' of Governor William Weld's Medical Information. Soc Sci Res Netw 2012:19.
57. Emam KE, Dankar FK, Vaillancourt R, Roffey T, Lysyk M. Evaluating the Risk of Re-identification of Patients from Hospital Prescription Records. Can J Hosp Pharm 2009 Jul;62(4):307-319 [FREE Full text] [Medline: 22478909]
58. Kulynych J, Greely HT. Clinical genomics, big data, and electronic medical records: reconciling patient rights with research when privacy and science collide. J Law Biosci 2017 Apr;4(1):94-132 [FREE Full text] [doi: 10.1093/jlb/lsw061] [Medline: 28852559]
59. Mascalzoni D, Paradiso A, Hansson M. Rare disease research: Breaking the privacy barrier. Appl Transl Genom 2014 Jun 01;3(2):23-29 [FREE Full text] [doi: 10.1016/j.atg.2014.04.003] [Medline: 27275410]
60. Zaaijer S, Gordon A, Speyer D, Piccone R, Groen S, Erlich Y. Rapid re-identification of human samples using portable DNA sequencing. eLife Internet 2019:6. [doi: 10.7554/elife.27798]
61. Ienca M, Ferretti A, Hurst S, Puhan M, Lovis C, Vayena E. Considerations for ethics review of big data health research: A scoping review. PLoS ONE 2018 Oct 11;13(10):e0204937. [doi: 10.1371/journal.pone.0204937]
62. Loukides G, Denny JC, Malin B. The disclosure of diagnosis codes can breach research participants' privacy. J Am Med Inform Assoc 2010;17(3):322-327 [FREE Full text] [doi: 10.1136/jamia.2009.002725] [Medline: 20442151]
63. Malin B, Benitez K, Masys D. Never too old for anonymity: a statistical standard for demographic data sharing via the HIPAA Privacy Rule. J Am Med Inform Assoc 2011;18(1):3-10 [FREE Full text] [doi: 10.1136/jamia.2010.004622] [Medline: 21169618]
64. El Emam K, Moher E. Privacy and anonymity challenges when collecting data for public health purposes. J Law Med Ethics 2013 Mar;41 Suppl 1:37-41. [doi: 10.1111/jlme.12036] [Medline: 23590738]
65. Thorogood A, Zawati M. International Guidelines for Privacy in Genomic Biobanking (or the Unexpected Virtue of Pluralism). J Law Med Ethics Internet 2015;43:4. [doi: 10.1111/jlme.12312]
66. O'Neill L, Dexter F, Zhang N. The Risks to Patient Privacy from Publishing Data from Clinical Anesthesia Studies. Anesth Analg 2016 Dec;122(6):2017-2027. [doi: 10.1213/ANE.0000000000001331] [Medline: 27172145]
67. Househ M, Grainger R, Petersen C, Bamidis P, Merolli M. Balancing Between Privacy and Patient Needs for Health Information in the Age of Participatory Health and Social Media: A Scoping Review. Yearb Med Inform 2018 Aug;27(1):29-36 [FREE Full text] [doi: 10.1055/s-0038-1641197] [Medline: 29681040]
68. Chevrier R, Foufi V, Gaudet-Blavignac C, Robert A, Lovis C. Use and Understanding of Anonymization and De-Identification in the Biomedical Literature: Scoping Review. J Med Internet Res 2019 May 31;21(5):e13484. [doi: 10.2196/13484]
69. Dixon P. A Failure to "Do No Harm" -- India's Aadhaar biometric ID program and its inability to protect privacy in relation to measures in Europe and the U.S. Health Technol (Berl) 2017;7(4):539-567 [FREE Full text] [doi: 10.1007/s12553-017-0202-6] [Medline: 29308348]
70. Rengers JM, Heyse L, Otten S, Wittek RPM. "It's Not Always Possible to Live Your Life Openly or Honestly in the Same Way" - Workplace Inclusion of Lesbian and Gay Humanitarian Aid Workers in Doctors Without Borders. Front Psychol 2019 Feb 27;10:320 [FREE Full text] [doi: 10.3389/fpsyg.2019.00320] [Medline: 30873072]
71. Stangl AL, Lloyd JK, Brady LM, Holland CE, Baral S. A systematic review of interventions to reduce HIV-related stigma and discrimination from 2002 to 2013: how far have we come? J Int AIDS Soc 2013 Nov 13;16(3 Suppl 2):18734 [FREE Full text] [Medline: 24242268]
72. Bélisle-Pipon J, Vayena E, Green RC, Cohen IG. Genetic testing, insurance discrimination and medical research: what the United States can learn from peer countries. Nat Med 2019 Aug;25(8):1198-1204. [doi: 10.1038/s41591-019-0534-z] [Medline: 31388181]

73. Bardey D, De Donder P, Mantilla C. How is the trade-off between adverse selection and discrimination risk affected by genetic testing? Theory and experiment. J Health Econ 2019 Sep 30;68:102223. [doi: 10.1016/j.jhealeco.2019.102223] [Medline: 31581025]
74. Global Alliance for Genomics & Health. URL: https://www.ga4gh.org/ [accessed 2019-10-05]
75. EU GDPR.ORG. URL: https://eugdpr.org/ [accessed 2019-10-05]
76. JMIR Medical Informatics. E-collection 'Big Data' URL: https://medinform.jmir.org/themes/183
77. JMIR Medical Informatics. E-collection 'Decision Support for Health Professionals' URL: https://medinform.jmir.org/themes/186
78. Journal of Medical Internet Research. E-collection 'Artificial Intelligence (other than Clinical Decision Support)' URL: https://www.jmir.org/themes/797

Abbreviations

AI: artificial intelligence
CPOE: computerized provider order entry
GA4GH: Global Alliance for Genomics and Health
GDPR: General Data Protection Regulation
RECIST: Response Evaluation Criteria In Solid Tumors
SNOMED CT: Systematized Nomenclature of Medicine - Clinical Terms

Edited by G Eysenbach; submitted 09.10.19; peer-reviewed by A Benis, D Gunasekeran; comments to author 13.10.19; revised version received 18.10.19; accepted 20.10.19; published 08.11.19.

Please cite as:
Lovis C
Unlocking the Power of Artificial Intelligence and Big Data in Medicine
J Med Internet Res 2019;21(11):e16607
URL: https://www.jmir.org/2019/11/e16607
doi:10.2196/16607
PMID:31702565

Review

Smartphone Apps to Support Coordinated Specialty Care for Prodromal and Early Course Schizophrenia Disorders: Systematic Review

Erica Camacho[1], BS; Leonard Levin[1], MSLIS; John Torous[1], MBI, MD

Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, United States

Corresponding Author:
John Torous, MBI, MD
Beth Israel Deaconess Medical Center
Harvard Medical School
330 Brookline Ave
Boston, MA, 02215
United States
Phone: 1 6176676799
Email: jtorous@bidmc.harvard.edu

Abstract

Background: Demand for mental health services, especially for clinical high-risk and early psychosis, has increased, creating a need for new solutions to increase access to and quality of care. Smartphones and mobile technology are potential tools to support coordinated specialty care for early psychosis, given their potential to augment the six core roles of care: case management and team leadership, recovery-oriented psychotherapy, medication management, support for employment and education, coordination with primary care services, and family education and support. However, the services smartphones are actually offering specifically for coordinated specialty care and the level of evidence are unknown.

Objective: This study aimed to review the published literature on smartphone technology to enhance care for patients with prodromal and early course psychosis and schizophrenia and to analyze studies by type, aligned with coordinated specialty care domains.

Methods: A systematic literature search was conducted on August 16 and 17, 2019, using the PubMed, EMBASE, Web of Sciences, and PsycINFO electronic databases. The eligible studies were reviewed and screened based on inclusion and exclusion criteria.

Results: The search uncovered 388 unique results, of which 32 articles met the initial inclusion criteria; 21 eligible studies on 16 unique app platforms were identified. Feasibility studies showed a high user engagement and interest among patients, monitoring studies demonstrated a correlation between app assessments and clinical outcomes, and intervention studies indicated that these apps have the potential to advance care. Eighteen studies reported on app use for the case management roles of coordinated specialty care. No app studies focused on employment and education, coordination with primary care services, and family education and support.

Conclusions: Although the published literature on smartphone apps for prodromal and first-episode psychosis is small, it is growing exponentially and holds promise to augment both monitoring and interventions. Although the research results and protocols for app studies are not well aligned with all coordinated specialty care roles today, high rates of adoption and feasibility suggest the potential for future efforts. These results will be used to develop coordinated specialty care–specific app evaluation scales and toolkits.

(J Med Internet Res 2019;21(11):e16393) doi:10.2196/16393

KEYWORDS

smartphones; mobile phones; app; schizophrenia

Introduction

Psychotic disorders are among the most disabling disorders in all of medicine [1]. The costs to society are greater than nearly any other chronic health condition, and the burden to patients and family members is of the greatest magnitude [2-4]. These disorders generally surface during late adolescence or early adulthood [5]. Individuals at clinical high risk are at a greater risk of developing a psychotic disorder; this clinical state is distinguished by impaired functioning, decreased quality of life, and subthreshold psychotic symptoms [6]. Similarly, first-episode psychosis is characterized by a decline in social functioning, the onset of psychotic symptoms like unusual thoughts, hallucinations, and impaired cognitive abilities [7]. Early treatment for clinical high risk and first-episode psychosis is a global health priority that aims to prevent or mitigate the burdens of the illness [8,9]. In this paper, we explore the available evidence for smartphone technology to augment clinical high risk and first-episode psychosis care.

Coordinated specialty care is an evidence-based, recovery-oriented treatment program designed to transform outcomes in first episode psychosis by promoting shared decision making and creating individualized treatment plans. The major roles of care include case management and team leadership, recovery-oriented psychotherapy, medication management, employment and education support, coordination with primary care services, and family education and support [10]. In the United States, there are currently 236 coordinated specialty care programs offering care for first-episode psychosis. Scaling up coordinated specialty care to reach more patients and reduce the duration of untreated psychosis is the next step in expanding services to those with early course psychosis.

One means to augment coordinated specialty care is leveraging technology like smartphones [11]. The rapidly evolving literature on smartphone apps for psychosis spectrum illnesses and a recent review on these apps already suggests its many potential roles in supporting facets of coordinated specialty care [12]. Specific advances around digital phenotyping and just-in-time adaptive interventions hold especially unique promise. By automatically quantifying patients' treatment trajectories through digital phenotyping, that is, "the moment-by-moment quantification of the individual-level human phenotype in situ using data from personal digital devices" [13], mobile technology could uniquely help offer functional assessments of recovery and personalization of care. These technologies can also potentially help screen and identify those with clinical high risk and ensure appropriate referral of patients with first-episode psychosis to coordinated specialty care that would reduce the duration of untreated psychosis. Through symptom surveys, therapy-based coaching, peer-to-peer support, and medication reminders, smartphone apps have direct potential to support coordinated specialty care today with just-in-time adaptive interventions.

Patients with clinical high risk and first-episode psychosis are already using smartphones and technology today. A 2014 survey conducted on 67 individuals with first-episode psychosis revealed that 88% of the participants had access to phones [14]. Research on smartphone ownership among youth receiving early psychosis–related services suggests that like the rest of the population, people with early psychosis increasingly own devices; a 2015 study reported 81% ownership [3] and 2018 research from our team reported 85% ownership [4]. These studies showcase the feasibility of using technology in care services, as many at clinical high risk and in first-episode psychosis today are already actively using these devices. Given that the clinical needs of patients with early course psychosis are different from those who have had the disease for many years, key roles in coordinated specialty care are customized for early course illness. As these younger patients may use smartphones and technology differently, it is necessary to consider how mobile health may uniquely meet the needs of patients with early course psychosis.

This review paper presents the first of three steps to explore the potential of apps for coordinated specialty care. The next stage will be guided by this review and involves the creation of specialized app evaluation scales, toolkits, and implementation guidelines for coordinated specialty care that will focus on five domains: (1) barriers to application, (2) sustainability following initial implementation, (3) training requirements/burden, (4) necessary infrastructure required to support integration, and (5) sustainable and durability of the technology. The final stage will include eight coordinated specialty care site visits to gather feedback from multiple stakeholders, assess the current state of technology readiness and need, and offer recommendations for successful technology implementation.

There is a clear potential for smartphones and mobile technologies to augment clinical high risk and first-episode psychosis care [12]. In this systematic review, we explore the literature on clinical high risk and first-episode psychosis care, with the goal of assessing feasibility around monitoring and interventions, aligning apps with core roles of coordinated specialty care, and identifying emerging trends. Given the nascent and heterogeneous nature of this space, we respect that different technologies, study designs, populations, and outcomes will preclude formal analysis and permit narrative synthesis.

Methods

Overview

A PICO (P - adolescents between the ages of 13 and 26 years, I - mobile/smartphone apps, C - no technological intervention, O - management of prodromal and early course schizophrenia and psychosis) and searchable question was created around the use of mobile technologies in an adolescent population experiencing prodromal and early course schizophrenia and psychosis. Search strategies were developed by author LL, who is a health sciences librarian. LL translated the search strategies based upon each database's respective controlled vocabulary (Medical Subject Headings, Emtree, Web of Science topic terms, Thesaurus of Psychological Index Terms for PsycINFO) and command language, using additional free-text terms when appropriate. Major concepts searched were *Schizophrenia* and *Psychotic Disorders* combined with *Early Diagnosis* and then joined to a list of applicable terms and synonyms indicating the use of mobile or wireless devices, smartphones, or apps.

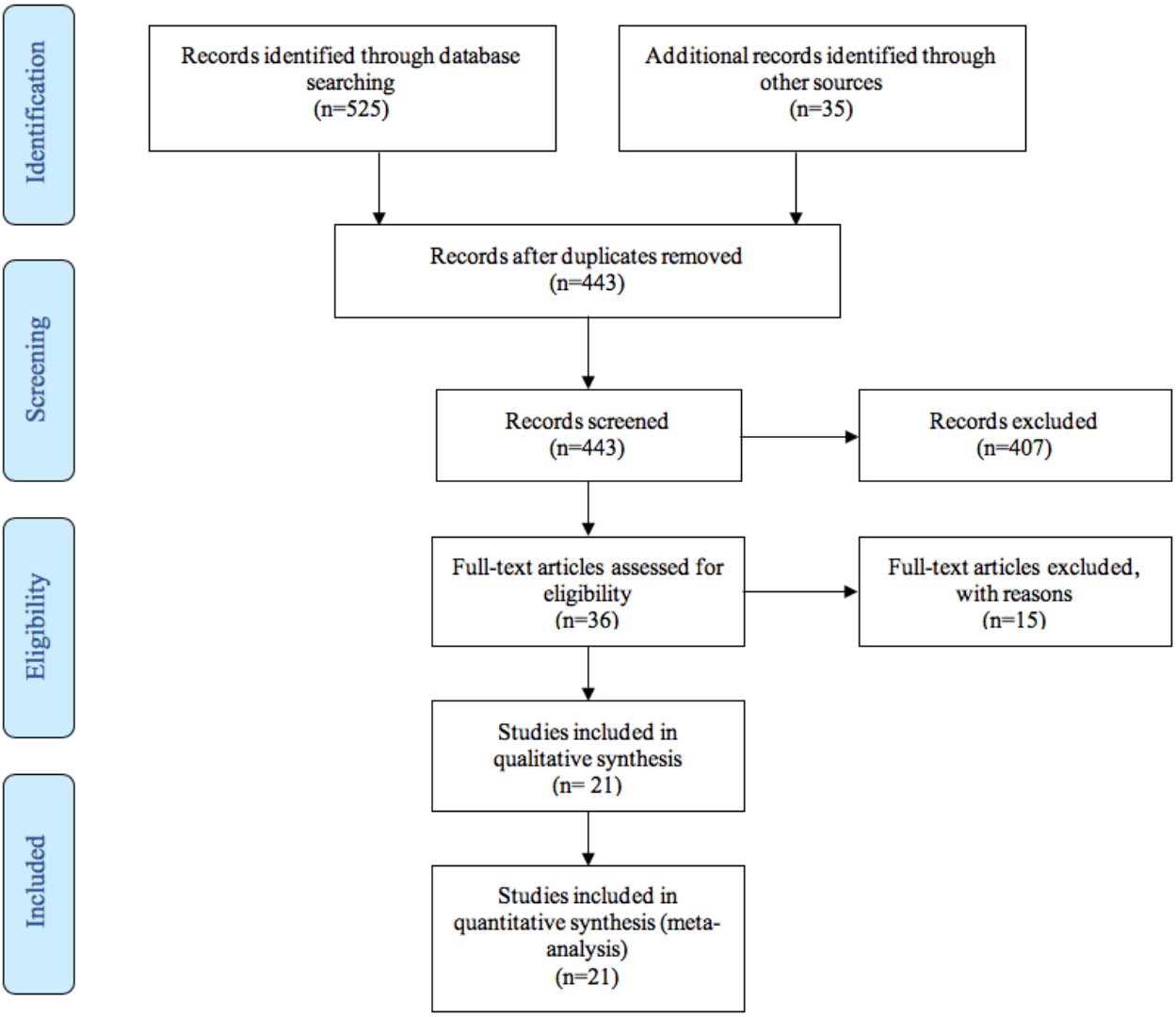
PubMed, EMBASE, Web of Sciences, and PsycINFO searches were conducted on August 16 and 17, 2019. Criteria included material written in English; age range of 13 through 26 years (adolescent and young adult); and material published from the beginning of 2008, the year in which a medical category was first introduced into the Apple platform app store.

A total of 525 articles were originally discovered. Following the elimination of duplicates, a total of 388 references were identified. Two authors (EC and JT) reviewed each citation/abstract using the Covidence (Melbourne, Australia) systematic review management tool. Inclusion criteria comprised publication date of January 1, 2008, or later; English language; adolescents between the ages of 13 and 26 years exhibiting prodromal, early course, or first-episode schizophrenia or at a high risk for clinical psychosis; and usage or the availability of a mobile/wireless/smartphone app. The following papers were excluded: articles not published in English, review articles, conference papers or poster abstracts, and any technology not delivered via a mobile device such as a desktop computer program or website.

Study Selection

Review of the 388 references was conducted by reviewing each abstract. Each reviewer appraised each article independently. If the abstract was not present or unclear, the full text of the article was retrieved. Following the initial screening, 32 articles met the initial inclusion criteria. Using bibliographies and cited-by references in these papers, hand searching was conducted. In addition, a set of text words culled from these papers was searched in Google Scholar in order to identify any additional grey literature. This process identified 35 additional articles. After full-text review of this final set as well as the original 32 articles, 21 articles met the full inclusion criteria and are reviewed here (Figure 1).

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram.



Results

Clinical High Risk

Protocol

The Robin Z app aims to reduce at-risk symptomatology and comorbid diagnosis while improving functioning, self-efficacy, and quality of life by offering elements of cognitive behavioral therapy, systematic therapy, and self-assessment between therapy sessions. Robin Z features include information and suggestions on coping with their symptoms, medication reminders, crisis intervention planning created with a therapist, weekly goals, and a library of positive reinforcements. Traber-Walker et al [15] plan to evaluate the app in a controlled study with 30 participants aged 14-18 years.

Monitoring

The ClinTouch app, one of the first in the psychosis space, collects ecological momentary assessment data from smartphones. The 2012 ClinTouch study lasted for 6 days and split patients into 3 groups: (1) patients in partial or full remission, (2) patients acutely psychotic, and (3) patients at ultra-high risk. Of the 44 participants entered in the study, 6 acute and 2 remitted patients did not complete the minimum number of diary entries and were excluded from analysis. No ultra-high risk participants were excluded. The ultra-high risk group completed more surveys (31.1) during the study than the remitted (29.5) or acute (28.5) group [16]. The mean age of the 12 ultra-high risk participants was 22 years, and 10 of 12 were male. The results suggest that smartphone ecological momentary assessment has high internal consistency and sensitivity to change in all groups, including ultra-high risk.

Intervention

MOMENTUM seeks to enrich social functioning in patients at clinical high risk via incorporation of the self-determination theory of motivation and a strength- and mindfulness-based approach. In a 2-month study [17] with 13 clinical high-risk patients (mean age 20.3) and in the 2019 study [18], the program could be accessed on a computer, tablet, or smartphone. The study represents the first efforts to offer interventions for high clinical risk. Results demonstrated a statistically significant improvement in social functioning for all participants, measured on the Global Functioning Scale, and 42% of participants showed improvements in the Satisfaction With Life scale [17].

First-Episode Psychosis

Protocol

Actissist is a digital intervention that focuses on five major domains associated with relapse in early psychosis: (1) perceived criticism, (2) socialization, (3) cannabis use, (4) paranoia, and (5) auditory verbal hallucinations. Smartphone notifications are sent at "3 psuedorandomized time points per day, 6 days a week between the hours 10am to 10pm" to encourage the participant to access the app [19]. This 2015 study explored feasibility and acceptability in patients with first-episode psychosis and demonstrates the app's ability to reduce psychotic symptoms and cannabis misuse while enhancing their quality of life. Outcomes of this now completed study are illustrated under the intervention section for first-episode psychosis.

The TechCare study [20] will blend experiential sampling methodology, which examines context and flow of "thoughts, feelings, and events," with intelligent real-time therapy, which delivers psychological interventions such as cognitive behavioral therapy. In this three-phase study, the first phase will involve collecting qualitative data through focus groups, which will be used to refine the intervention for the next phase. This second phase will evaluate the acceptability of TechCare. The third and final phase will be a feasibility trial, where case coordinators will help create personalized interventions by customizing the app and linking ecological momentary assessment responses to treatment options including a crisis plan [20].

HORYZONS is an internet-based platform that was developed to prevent relapse and improve social functioning in users. The platform aims to support long-term recovery in first-episode psychosis by offering therapy modules and features that promote behavior change and social networking. This study protocol adapts earlier foundational work [21] from Australia to now meet the cultural needs of patients with first-episode psychosis in Canada [22]. The program can be accessed via a smartphone, and the study goals are to gain insight on the acceptability of the platform and discover recommendations on how to improve it.

HORYZONS will be utilized for a randomized controlled trial in Melbourne to test whether its use for first-episode psychosis can enhance social functioning and reduce hospitalizations when compared to treatment as usual. Alvarez-Jimenez et al [23] plan to recruit 170 participants with first-episode psychosis, aged 16-27 years, to partake in the 18-month study. The study will utilize ecological momentary assessment to assess positive effect, negative effect, and social isolation via surveys eight times a day, within a 12-hour window. It also features "online pathways" designed to improve participant self-efficacy such as "fostering positive emotions" and "identifying and exercising personal strengths" and is supported by both peer and clinician moderators. Expert moderators, that is, clinicians, will also be able to customize the app and treatment approaches to each patient's needs.

When planning the Psychotherapy app, Barbeito et al [24] wrote a protocol for a focus group, to gain insight into the opinions of young people with psychosis in Spain [24]. The app, comprised of five modules outlined in Table 1, will be studied to assess its ability to decrease the number of relapses and rehospitalizations in 50 patients with first-episode psychosis, aged 14-19 years, when compared to treatment as usual. Features of the app will be supported by first-episode psychosis users such as the contact wall, while others will be moderated and customized by clinicians.

Table 1. Studies of smartphone apps for people with first-episode psychosis or at clinical high risk.

Parameter and intervention/app	Author, year	Main findings
CHR[a] - Protocol		
Robin Z	Traber-Walker et al, 2019 [15]	The Robin Z app offers support between therapy sessions with the goal of improving the daily functioning of people in CHR states. The study goal is to decrease at-risk symptoms like delusions, depression, and hallucinations following usage of the app in addition to TAU[b].
CHR - Monitoring		
ClinTouch	Palmier-Claus et al, 2012 [16]	The ClinTouch app offers symptom assessments for CHR patients and provides clinical information to their providers remotely. The study concluded that smartphones apps are a valid method for symptom management as seen from high participant compliance rates.
CHR - Intervention		
MOMENTUM	Alvarez-Jimenez et al, 2018 [17]	The MOMENTUM app is designed to improve the self-efficacy of people at ultra-high risk for psychosis by helping participants focus on their strengths, practice mindfulness, and connect with one another. Results demonstrated improvements in social functioning and wellness as well as high engagement and satisfaction with the app.
FEP[c] - Protocol		
Actissist	Bucci et al, 2015 [19]	Actissist is an intervention that focuses on five domains that are associated with early psychosis relapse. The study will compare it to ClinTouch, a symptom monitoring app, plus TAU.
TechCare	Husain et al, 2016 [20]	TechCare blends experiential sampling methodology and intelligent real-time therapy to provide participants with both assessments and interventions. The study collected user feedback to refine the intervention and to test the feasibility of the app.
HORYZONS	Lal et al, 2018 [22]	HORYZONS is a Web platform, accessible via a smartphone, that is capturing feedback from Canadian youth on the framework with the purpose of adapting the program to better serve those with FEP.
HORYZONS	Alvarez-Jimenez et al, 2019 [23]	HORYZONS will utilize a smartphone ecological momentary assessment tool to deliver surveys and interactive therapy content to FEP participants with a focus on improving social functioning.
Psychotherapy	Barbeito et al, 2019 [24]	The Psychotherapy app study will investigate whether five modules in the app may minimize relapse and hospitalization in FEP compared to TAU. The modules include psychoeducation, recognition of symptoms and prevention of relapses, problem solving, mindfulness, and a contact wall.
MOMENTUM	Vitger et al, 2019 [18]	MOMENTUM is converted into a smartphone app to be utilized for FEP. The emphasis of this study will be on improving shared decision making between patients and their carers.
FEP - Usability/feasibility		
Unnamed app	Smelror et al, 2019 [25]	Smelror et al conducted an exploratory study for the use of an app to assist patients with early onset psychosis to manage their auditory verbal hallucinations.
PRIME	Schlosser et al, 2016 [26]	PRIME is an intervention app for FEP that provides patients with goal-setting tools, CBT[d]-based coaching from a clinician, and social networking opportunities with their peers. The study showed high engagement rates, 100% retention, and high user satisfaction to conclude that the app is feasible and acceptable.
Heal Your Mind	Kim et al, 2018 [27]	The Heal Your Mind app offers case management and symptom monitoring for young people with early psychosis. The surveys collected showed that a majority of participants used at least 5 of the 6 modules, felt the app was easy to use, and expressed satisfaction with the tool.
+Connect	Lim et al, 2019 [28]	+Connect is an intervention app designed to target loneliness in youth with early psychosis. The study outcomes showed a decrease on the University of California Loneliness scale.
ACT-DL[e]	Vaessen et al, 2019 [29]	The ACT-DL app utilizes acceptance and commitment therapy to help patients with early psychosis improve negative symptoms. The study showed that participants found the app to be a useful tool to solidify knowledge gained from weekly therapy sessions.

Parameter and intervention/app	Author, year	Main findings
RealLife Exp	Kuman et al, 2018 [30]	The RealLife Exp app is used alongside a Web-based dashboard to help early psychosis patients with symptom monitoring. Study outcomes indicate that participants are moderately responsive to daily and weekly assessments.
FEP - Monitoring		
CrossCheck	Ben-Zeev et al, 2017 [31]	The CrossCheck app collects ecological momentary assessments, device use data, and passive data like geolocation to predict relapse in people with psychosis. Study outcomes indicate that digital indicators of relapse are not the same for every individual experiencing psychosis.
mindLAMP	Wisniewski et al,2019 [32]	The mindLAMP app also collects ecological momentary assessments, device use data, and passive data like geolocation. Study outcomes indicate that digital markers can help inform changes in clinical care.
Ginger.io	Niendam et al, 2018 [33]	Ginger.io is a symptom-monitoring app for individuals with early psychosis, which collects survey responses and passive data like distance travelled and phone calls. The study showed that the app is easy to use and a willingness of patients to continue incorporating apps into the patient's treatment plan.
ClinTouch	Cella et al, 2019 [34]	ClinTouch is used alongside a wearable device to draw conclusions on whether there is a connection between distressing psychosis symptoms and physiological responses. The study outcomes showcase increased electrodermal activity when experiencing hallucinations or delusions, but no association between symptoms and heart rate variability.
FEP - Intervention		
PRIME	Schlosser et al, 2018 [35]	The PRIME app seeks to improve motivation in FEP. The study findings show improvements in reward learning, anticipated pleasure, and effort expenditure for the PRIME group as compared to the waitlist.
Actissist	Bucci et al, 2018 [36]	The Actissist app plus TAU showed greater and more sustained treatment effects and benefits as compared to a symptom monitoring app plus TAU. A majority of participants from the Actissist arm submitted at least half of their data entries, and all members of this arm were retained.

[a]CHR: clinical high risk.

[b]TAU: treatment as usual.

[c]FEP: first-episode psychosis.

[d]CBT: cognitive behavioral therapy.

[e]ACT-DL: acceptance and commitment therapy in daily life.

The 2019 MOMENTUM protocol aims to assess how app use in patients with first-episode psychosis can augment shared decision making with their providers. With a planned enrollment at 260 participants aged 18-35 years, it represents the largest sample size planned to date [18]. The app will be used by the participant to explore and evaluate aspects of their daily life such as sleep and stress and then share those data with the providers. By increasing the patients' knowledge of and comfort with their symptoms in parallel with the providers' increased awareness of daily changes, the study hypothesizes that app use will facilitate shared decision making in first-episode psychosis.

Usability/Feasibility

Smelror et al [25] conducted an exploratory study for 7 days by using an app to assist patients with early onset psychosis to cope with their auditory verbal hallucinations. Of the four patients contacted to participate in the study, one refused and one was excluded from the statistical analyses due to poor compliance. Of note, one participant reported discomfort with digital monitoring and noted feeling watched, monitored, and insecure. The study concluded that adolescents with early onset psychosis are willing to use apps to self-assess their symptoms and apps are a viable option for this population.

The PRIME app is an intervention tool, supported by peers and coaches, which aims to improve quality of life and negative symptoms like motivation in first-episode psychosis. In this two-phase study, the first 10 enrolled participants (mean age 23.4) used the app to determine feasibility. Based on phase one feedback, the second phase was an ongoing randomized controlled trial with the first 10 participants (mean age 23.3) testing the effectiveness of the app's iterative design process. Over 12 weeks, participants logged in about every other day and completed 1.5 daily challenges a week, such as working out for 30 minutes. Retention was high at 100%, and the average level of satisfaction for all participants was 8/10 [26]. The study concluded that PRIME is a feasible and highly acceptable intervention tool for first-episode psychosis.

The Heal Your Mind app provides cognitive behavioral therapy–based case management and symptom monitoring for young people with first-episode psychosis through six modules: (1) thought record, (2) symptom record, (3) daily life record, (4) official notices, (5) communication, and (6) scales [27]. This feasibility study involved 33 early intervention service users (mean age 25.6). Usage reports were high, with 41.7% of participants completing all 6 modules [27]. The most frequently

used, liked, and perceived helpful feature was communication with the case manager. High satisfaction was reported in about 80% of participants and 46% felt the app was useful for monitoring their symptoms. One participant reported feelings of stress while using the app. The study concluded that the app is feasible and useful for young people experiencing early psychosis.

The +Connect app delivers positive psychology interventions via three different video types: peer videos, expert videos, and actor videos. Videos are offered in response to real-time mood tracking. The 6-week study required 12 participants (mean age 20.5) to complete the daily activities on the app for at least 70% of the duration. Both user engagement and satisfaction were high at 95.47% and 90%, respectively [28]. Conclusions of the study were that the app is highly acceptable, feasible, and usable.

The acceptance and commitment therapy in daily life (ACT-DL) study [29] utilized a subset of participants to determine the feasibility of an ongoing randomized controlled trial, which plans to enroll 150 participants into the program. The ACT-DL app offers ACT toward improving negative symptoms. In this clinical hybrid study, participants meet with ACT therapists weekly and then apply the lessons learned on the app platform for a minimum of 3 consecutive days. Participants are prompted to answer a short questionnaire on their mood and symptoms. The app then elicits an ACT metaphor or exercise based on the module of the week. The 16 participants to date found both the ACT therapy sessions and the home exercises useful. They also reported that the app helped them implement ACT into their daily life and helped bring awareness to their emotions [29].

The RealLife Exp app provides symptom monitoring for patients with early psychosis through daily and weekly smartphone surveys. The 5-month study involved 61 participants with early psychosis interacting with the app and 20 providers who connected through the online dashboard. Of the 41 participants who completed the study, 66% reported a willingness to continue using RealLife Exp as part of their treatment service, while 12% reported a lack of interest in using the app [30]. A majority of participants suggested improvements to the app, such as technical functionality, product enhancements, and changes to the survey.

Monitoring

A case series paper [31] of the CrossCheck app featured two patients with first-episode psychosis, aged 19 years, likely meeting. The CrossCheck app combines ecological momentary assessment with digital phenotyping (smartphone-based) geolocation, speech frequency and duration, and physical activity. The app was preinstalled on smartphones with an unlimited data plan provided to participants in the study. With a 12-month data-collection period, it represents the longest reported use of a mobile health smartphone app for people with early psychosis. In the first case, a 19-year-old African-American male with schizophrenia had ecological momentary assessment data for 91% of the 125 days before hospitalization and audio sensor data for 87% of the days. The audio sensor data showed a decrease in speech frequency and duration over 50 days and then an increase over 70 days, followed by a spike in both before hospitalization. In the second case, a 19-year-old Hispanic American-Indian female with schizophrenia had ecological momentary assessment data for 91% of the 190 days prior to hospitalization and device use data for 93% of the days. Device use data showed that the patient's phone was unlocked between midnight and 6 AM during the 60 days before hospitalization, which was unusual compared to her first 100 days of data [31]. Similarly, our team wrote a case series exploring the role of passive data monitoring in patients with psychosis, including first-episode psychosis, using the mindLAMP app [32].

Ginger.io offers another platform for both ecological momentary assessment and digital phenotyping. In this clinical hybrid study, including 64 participants with recent-onset psychosis and 12 with clinical high risk, the app collected active data through self-report surveys and digital phenotyping including phone calls, messages, and distance traveled [33]. As a clinical hybrid, the study involved monthly in-person psychosocial assessments with the research team. A total of 97% percent of the 60 participants who completed satisfaction surveys found the app easy to use, and 83% were open to continued use of the app in their treatment plan [33]. The authors report that smartphone assessments of symptoms were comparable to the Brief Psychosis Rating Scale conducted at the monthly clinician interviews.

The 2019 ClinTouch study [34] was a 10-day observational cohort study exploring the association between symptoms of psychosis, such as hallucinations and delusions, to physiological responses, such as heart rate variability and electrodermal activity. The study used a wearable, E4, to measure physiological changes and used the mobile app to conduct the self-assessment. Smartphone notifications were sent at "4 pseudo-randomized time points per day between the hours 11am to 9pm" to prompt participants to rate their symptoms [34]. One participant dropped out due to a non–research-related reason. Of the 14 participants who completed the study, 76% of surveys were completed [34]. As distressing hallucinations and delusions were reported, there was a significant increase in electrodermal activity. No significant association was found between these symptoms and heart rate variability [34]. Participants found the app to be easy to use, nondisruptive, and enjoyable overall.

Intervention

The 2018 PRIME study [35] conducted a 12-week randomized controlled trial to test the app's ability to enhance motivational impairments in first-episode psychosis. Of the 43 participants recruited (mean age: 24.3), 22 were in the PRIME group, 21 were in the treatment as usual/waitlist group, 5 dropped out, and 6 did not complete the follow-up. To assess changes in motivated behavior, the trial modified the Trust Task to determine three aspects of motivation: reward learning, anticipated pleasure, and effort expenditure. Only PRIME participants were given the Trust Task. These participants showed a greater increase in anticipated pleasure, effort expended to increase the likelihood of future social interactions with positive outcomes, and learning from positive outcomes from baseline to 12 weeks compared to waitlist [35]. The retention rate for the treatment was 74%, and mean satisfaction with PRIME was 8.21/10. Participants noted that the app helped

them feel less alone, more hopeful, more connected, and less helpless. The ability to directly message their coaches was the most popular feature, while the ability to track mood was the least popular.

The 2018 Actissist proof-of-concept study [36] involved a 12-week randomized controlled trial for 36 participants. The study had a 2:1 ratio for the experimental group Actissist plus treatment as usual (n=24, mean age 20 years) to the control group ClinTouch plus treatment as usual (n=12, mean age 18 years). The study set a target criterion of half of the participants in each arm to submit at least half of the data entries. The Actissist arm met the target criterion (63%), while ClinTouch did not (42%) [36]. In addition, 75% of Actissist users engaged with the app at least once a day. The most popular prompted and unprompted domain was voices followed by suspicious thoughts. The Actissist arm showed larger treatment effects and greater benefits than symptom monitoring plus treatment as usual in the immediate posttreatment assessment and the 22-week follow-up.

Discussion

Principal Findings

In this systematic review, we examined the published literature on smartphone apps for both prodromal and first episode psychosis and found 21 studies. Seven papers were protocols, six were feasibility studies, five were monitoring/validity studies, and three were interventions. The heterogeneous outcomes of the 21 studies and 16 unique app platforms utilized preclude formal analysis, and data on engagement, adherence, and feasibility varied greatly, as outlined in the narrative results. Overall, feasibility studies were positive and reported high engagement and interest among patients, monitoring studies showed that apps captured outcomes correlated to and informative of clinical outcomes, and intervention studies demonstrated that these apps have the potential to advance care.

Every study aligned with at least one core role of coordinated specialty care but none aligned with all. Of the 21 studies, 18 incorporated case manage mentor team leadership, 12 involved recovery-oriented psychotherapy, and 7 mentioned medication management. A majority of these apps offer case management through crisis planning and symptom monitoring. This finding aligns with recent reviews [12]. Recovery-oriented psychotherapy in these apps was offered by both clinicians and peers. None of the studies reported provision of support for employment and education, coordination with primary care services, or family education and support, as seen in Textbox 1. This does not mean that apps cannot support coordinated specialty care; instead of a single app offering resources to meet all needs, a toolkit approach of several apps may be more feasible today. Although many apps for adult psychosis exist and could offer services for individual roles of coordinated specialty care, it is unclear if such apps designed and assessed in adult patients would equally benefit youth.

Many of the apps reviewed offered social and peer support as their main feature or intervention [17,22-24,26,35]. For example, HORYZONS incorporates peer moderators in the “café” feature where youth with mental illness can share their lived experiences to help participants navigate issues they may be facing. PRIME users expressed positive feedback from the social support they received from peers and their coaches. Other apps provide space for peer interaction, such as MOMENTUM and the Psychotherapy app. The Psychotherapy app uniquely provides two avenues for peer contact, including both more formal peer support and less formal peer interactions. One implementation challenge in moving promising research toward peer-to-peer forums is the need for moderation and safety mentoring, which can become difficult at scale.

Clinicians involved with these smartphone apps can have a variety of roles, such as moderating social interactions, reviewing patient clinical status, and customizing patient interventions. For example, through the Psychotherapy app and HORYZONS, clinicians can review user data to create personalized content suggestions. Ginger.io alerted clinicians when surveys were not completed for more than 3 days in a row, when passive data were not being collected, or when participant responses were considered clinically significant. The MOMENTUM study hypothesizes that an increased awareness of daily changes in the patient will better equip both parties for shared decision making [18]. These apps utilize their programs as a clinical hybrid to better fit the patient’s needs. Implementation of such a hybrid model of care will involve helping clinicians optimally work with technology, a focus of future efforts of this ongoing initiative.

The potential of smartphone apps for care is, in part, driven by their scalability, and our study results support the potential for sharing and reusing apps. The HORYZONS trial that adapted the Australian app to meet the treatment needs of youth in Canada, offers an example of global accessibility and collaboration possible through digital tools. The ClinTouch app has been used not only for feasibility studies of symptom reporting and physiological markers, but also for a control app in intervention studies, highlighting the multiple uses of these apps. It may be possible to use apps not directly targeting clinical high risk and first-episode psychosis; such a toolkit will also be the next focus of this initiative.

Textbox 1. Coordinated specialty care checklist.

Case management and team leadership: n=18
Recovery-oriented psychotherapy: n=13
Medication management: n=7
Supported employment and education: n=0
Coordination with primary care services: n=0
Family education and support: n=0

Although many studies focused on innovations in monitoring patients or interventions, few offered both. Results from numerous studies suggest that it is possible to capture self-reported symptoms via apps that are comparable to clinical assessments, and the 2019 ClinTouch study demonstrates how this can even be linked to physiology. Although the CrossCheck studies demonstrated the potential utility of app monitoring via digital phenotyping to predict relapse, no studies instigated the predictive validity of app monitoring. The TechCare study offers an example of a combined approach as well as how clinicians and smartphone apps can work symbiotically. The intervention portion of the app is tailored to the participant's survey responses and the specific delusions they are experiencing. If the patient is in a low mood or paranoid state for a prolonged period, the agreed-upon crisis plan is deployed. Tailoring interventions based on digital phenotyping data today would be more speculative, given the current lack of strong data, suggesting a reliable clinical interpretation of those data.

The growing interest in smartphone apps for prodromal and first-episode schizophrenia is reflected in the high number of recently published protocol papers we identified. These newly planned studies intend to expand the size and duration of completed ones, with a mean of 29.8 participants in completed studies compared to a targeted mean of 84.1 participants from protocols, as seen in Table 2. Future studies aim to increase the study duration by almost double to a mean duration of 5.7 months as compared to completed studies with a mean duration of 2.9 months. Not all planned, or even currently underway, studies published protocols; therefore, these results must be interpreted with some caution, although they offer insight into designing and powering the next wave of relevant studies.

The results of this review also suggest the low risk of harm in using app tools in patients with prodromal and first-episode schizophrenia. No adverse events were reported in any study, although some participants reported higher stress levels due to the fear of being watched or monitored [25]. This is likely not unique to their conditions, as across the general population, there is a small percent who are also not comfortable using technology and feel concerned about digital monitoring [37]. Some apps reported on security measures in place to protect privacy such as the Robin Z app, which requires fingerprint scanning or a pin code to login, but this was not reported consistently across studies.

Like all studies, this review has several limitations. First, no search term can identify all relevant papers on this topic and many publications are hard to recognize, given the varied nosology around prodromal, first-episode, and smartphone tools. Second, we used clinical judgment in including some studies such as the case report on CrossCheck, as not every paper directly identifies its population as first episode. Third, more apps and studies may exist that have not published protocols or data papers, meaning that our results are influenced by publication bias. We aimed to minimize these limitations by working with a librarian to build the search term and conduct the search.

Table 2. Summary metrics for studies on smartphone apps for people with first-episode psychosis or at clinical high risk.

Metric	Value	Studies
Studies in the United States, n (%)	6 (29)	21 studies [15-20,22-36]
Participants enrolled in completed studies, mean	29.8	12 studies [16,17,25-30,33-36]
Participants planned to be enrolled, mean	84.1	12 studies [16,17,25-30,33-36]
Participants dropped out from completed studies, mean	3.4	12 studies [16,17,25-30,33-36]
Age in completed studies (years), mean	21.5	11 studies [16,17,25-28,30,33-36]
Completed studies with male as the majority gender, n (%)	7 (70)	10 studies [16,17,26-28,30,33-36]
Duration of completed studies (months), mean	2.9	12 studies [16,17,25-30,33-36]
Duration of planned studies (months), mean	5.7	7 studies [15,19,20,22-18]

Conclusions

Results of this systematic review suggest that while the published research on smartphone apps for prodromal and first-episode psychosis is nascent, it is rapidly expanding and holds the potential to improve both monitoring and interventions. Although we did not find a single app that aims to fulfill all key roles of coordinated specialty care, these results are useful for understanding the current impact of mobile technology for care and future trends. The next steps in our work supporting such technology for coordinated specialty care will involve creating an evaluation and implementation framework that will help programs select and utilize a toolkit composed of several apps. By assessing these frameworks at coordinated specialty care sites across the country, we will learn which digital tools and implementations may best help sites augment care for patients with early course psychosis.

Acknowledgments

This work was supported by a contract from the National Institute of Mental Health (75N95019P00434).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Inclusion and exclusion criteria for Table 2.
[PDF File (Adobe PDF File), 52 KB - jmir_v21i11e16393_app1.pdf]

References

1. Dixon L. What It Will Take to Make Coordinated Specialty Care Available to Anyone Experiencing Early Schizophrenia. JAMA Psychiatry 2017 Jan 01;74(1):7. [doi: 10.1001/jamapsychiatry.2016.2665]
2. Savill M, Skymba HV, Ragland JD, Niendam T, Loewy RL, Lesh TA, et al. Acceptability of Psychosis Screening and Factors Affecting Its Implementation: Interviews With Community Health Care Providers. PS 2018 Jun;69(6):689-695. [doi: 10.1176/appi.ps.201700392]
3. Lal S, Dell'Elce J, Tucci N, Fuhrer R, Tamblyn R, Malla A. Preferences of Young Adults With First-Episode Psychosis for Receiving Specialized Mental Health Services Using Technology: A Survey Study. JMIR Mental Health 2015 May 20;2(2):e18. [doi: 10.2196/mental.4400]
4. Haller C, Padmanabhan J, Lizano P, Torous J, Keshavan M. Recent advances in understanding schizophrenia. F1000Prime Rep 2014;6:57 [FREE Full text] [doi: 10.12703/P6-57] [Medline: 25184047]
5. Gogtay N, Vyas NS, Testa R, Wood SJ, Pantelis C. Age of Onset of Schizophrenia: Perspectives From Structural Neuroimaging Studies. Schizophrenia Bulletin 2011 Apr 19;37(3):504-513. [doi: 10.1093/schbul/sbr030]
6. Fusar-Poli P, Rocchetti M, Sardella A, Avila A, Brandizzi M, Caverzasi E, et al. Disorder, not just state of risk: Meta-analysis of functioning and quality of life in people at high risk of psychosis. Br J Psychiatry 2018 Jan 02;207(3):198-206. [doi: 10.1192/bjp.bp.114.157115]
7. Greenstein L. NAMI: National Alliance on Mental Health. 2017. How Should We Be Treating First-Episode Psychosis? URL: https://www.nami.org/Blogs/NAMI-Blog/March-2017/How-Should-We-Be-Treating-First-Episode-Psychosis
8. Miller TJ, McGlashan TH, Rosen JL, Somjee L, Markovich PJ, Stein K, et al. Prospective Diagnosis of the Initial Prodrome for Schizophrenia Based on the Structured Interview for Prodromal Syndromes: Preliminary Evidence of Interrater Reliability and Predictive Validity. AJP 2002 May;159(5):863-865. [doi: 10.1176/appi.ajp.159.5.863]
9. McFarlane WR. Prevention of the first episode of psychosis. Psychiatr Clin North Am 2011 Mar;34(1):95-107 [FREE Full text] [doi: 10.1016/j.psc.2010.11.012] [Medline: 21333842]
10. Coordinated Specialty Care Fact Sheet Checklist. NIMH: National Institute of Mental Health. Coordinated Specialty Care Fact Sheet and Checklist URL: https://www.nimh.nih.gov/health/publications/raise-fact-sheet-coordinated-specialty-care/index.shtml [accessed 2019-11-01]
11. Torous J, Woodyatt J, Keshavan M, Tully LM. A new hope for early psychosis care: the evolving landscape of digital care tools. Br J Psychiatry 2019 Feb 11;214(5):269-272. [doi: 10.1192/bjp.2019.8]
12. Bell IH, Alvarez-Jimenez M. Digital Technology to Enhance Clinical Care of Early Psychosis. Curr Treat Options Psych 2019 Jul 11;6(3):256-270. [doi: 10.1007/s40501-019-00182-y]
13. Torous J, Keshavan M. A new window into psychosis: The rise digital phenotyping, smartphone assessment, and mobile monitoring. Schizophr Res 2018 Jul;197:67-68. [doi: 10.1016/j.schres.2018.01.005] [Medline: 29338959]
14. Lal S, Dell'Elce J, Malla AK. Technology Access and Use Among Young Adults With a First Episode of Psychosis. PS 2015 Jul;66(7):764-765. [doi: 10.1176/appi.ps.201400580]
15. Traber-Walker N, Gerstenberg M, Metzler S, Joris MR, Karr M, Studer N, et al. Evaluation of the Combined Treatment Approach "Robin" (Standardized Manual and Smartphone App) for Adolescents at Clinical High Risk for Psychosis. Front Psychiatry 2019;10:384 [FREE Full text] [doi: 10.3389/fpsyt.2019.00384] [Medline: 31244692]
16. Palmier-Claus JE, Ainsworth J, Machin M, Barrowclough C, Dunn G, Barkus E, et al. The feasibility and validity of ambulatory self-report of psychotic symptoms using a smartphone software application. BMC Psychiatry 2012 Oct 17;12(1). [doi: 10.1186/1471-244x-12-172]
17. Alvarez-Jimenez M, Gleeson J, Bendall S, Penn D, Yung A, Ryan R, et al. Enhancing social functioning in young people at Ultra High Risk (UHR) for psychosis: A pilot study of a novel strengths and mindfulness-based online social therapy. Schizophr Res 2018 Dec;202:369-377. [doi: 10.1016/j.schres.2018.07.022] [Medline: 30031616]
18. Vitger T, Austin SF, Petersen L, Tønder ES, Nordentoft M, Korsbek L. The Momentum trial: the efficacy of using a smartphone application to promote patient activation and support shared decision making in people with a diagnosis of schizophrenia in outpatient treatment settings: a randomized controlled single-blind trial. BMC Psychiatry 2019 Jun 17;19(1):185 [FREE Full text] [doi: 10.1186/s12888-019-2143-2] [Medline: 31208376]
19. Bucci S, Barrowclough C, Ainsworth J, Morris R, Berry K, Machin M, et al. Using mobile technology to deliver a cognitive behaviour therapy-informed intervention in early psychosis (Actissist): study protocol for a randomised controlled trial. Trials 2015 Sep 10;16:404 [FREE Full text] [doi: 10.1186/s13063-015-0943-3] [Medline: 26357943]

20. Husain N, Gire N, Kelly J, Duxbury J, McKeown M, Riley M, et al. TechCare: mobile assessment and therapy for psychosis - an intervention for clients in the Early Intervention Service: A feasibility study protocol. SAGE Open Med 2016 Oct 18;4:2050312116669613 [FREE Full text] [doi: 10.1177/2050312116669613] [Medline: 27790373]
21. Alvarez-Jimenez M, Bendall S, Lederman R, Wadley G, Chinnery G, Vargas S, et al. On the HORYZON: moderated online social therapy for long-term recovery in first episode psychosis. Schizophr Res 2013 Jan;143(1):143-149. [doi: 10.1016/j.schres.2012.10.009] [Medline: 23146146]
22. Lal S, Gleeson J, Malla A, Rivard L, Joober R, Chandrasena R, et al. Cultural and Contextual Adaptation of an eHealth Intervention for Youth Receiving Services for First-Episode Psychosis: Adaptation Framework and Protocol for Horyzons-Canada Phase 1. JMIR Res Protoc 2018 Apr 23;7(4):e100. [doi: 10.2196/resprot.8810]
23. Alvarez-Jimenez M, Bendall S, Koval P, Rice S, Cagliarini D, Valentine L, et al. HORYZONS trial: protocol for a randomised controlled trial of a moderated online social therapy to maintain treatment effects from first-episode psychosis services. BMJ Open 2019 Feb 19;9(2):e024104. [doi: 10.1136/bmjopen-2018-024104]
24. Barbeito S, Sánchez-Gutiérrez T, Mayoral M, Moreno M, Ríos-Aguilar S, Arango C, et al. Mobile App-Based Intervention for Adolescents With First-Episode Psychosis: Study Protocol for a Pilot Randomized Controlled Trial. Front Psychiatry 2019;10:27 [FREE Full text] [doi: 10.3389/fpsyt.2019.00027] [Medline: 30804818]
25. Smelror RE, Bless JJ, Hugdahl K, Agartz I. Feasibility and Acceptability of Using a Mobile Phone App for Characterizing Auditory Verbal Hallucinations in Adolescents With Early-Onset Psychosis: Exploratory Study. JMIR Form Res 2019 May 14;3(2):e13882 [FREE Full text] [doi: 10.2196/13882] [Medline: 31094321]
26. Schlosser D, Campellone T, Kim D, Truong B, Vergani S, Ward C, et al. Feasibility of PRIME: A Cognitive Neuroscience-Informed Mobile App Intervention to Enhance Motivated Behavior and Improve Quality of Life in Recent Onset Schizophrenia. JMIR Res Protoc 2016 Apr 28;5(2):e77. [doi: 10.2196/resprot.5450]
27. Kim S, Lee G, Yu H, Jung E, Lee J, Kim S, et al. Development and feasibility of smartphone application for cognitive-behavioural case management of individuals with early psychosis. Early Intervention in Psychiatry 2017 May 18;12(6):1087-1093. [doi: 10.1111/eip.12418]
28. Lim MH, Gleeson JFM, Rodebaugh TL, Eres R, Long KM, Casey K, et al. A pilot digital intervention targeting loneliness in young people with psychosis. Soc Psychiatry Psychiatr Epidemiol 2019 Mar 14. [doi: 10.1007/s00127-019-01681-2] [Medline: 30874828]
29. Vaessen T, Steinhart H, Batink T, Klippel A, Van Nierop M, Reininghaus U, et al. ACT in daily life in early psychosis: an ecological momentary intervention approach. Psychosis 2019 Mar 19;11(2):93-104. [doi: 10.1080/17522439.2019.1578401]
30. Kumar D, Tully LM, Iosif A, Zakskorn LN, Nye KE, Zia A, et al. A Mobile Health Platform for Clinical Monitoring in Early Psychosis: Implementation in Community-Based Outpatient Early Psychosis Care. JMIR Ment Health 2018 Feb 27;5(1):e15. [doi: 10.2196/mental.8551]
31. Ben-Zeev D, Brian R, Wang R, Wang W, Campbell AT, Aung MSH, et al. CrossCheck: Integrating self-report, behavioral sensing, and smartphone use to identify digital indicators of psychotic relapse. Psychiatr Rehabil J 2017 Sep;40(3):266-275 [FREE Full text] [doi: 10.1037/prj0000243] [Medline: 28368138]
32. Wisniewski H, Henson P, Torous J. Using a Smartphone App to Identify Clinically Relevant Behavior Trends Symptom Report, Cognition Scores, and Exercise Levels: A Case Series. Front Psychiatry 2019;10:652 [FREE Full text] [doi: 10.3389/fpsyt.2019.00652] [Medline: 31607960]
33. Niendam TA, Tully LM, Iosif A, Kumar D, Nye KE, Denton JC, et al. Enhancing early psychosis treatment using smartphone technology: A longitudinal feasibility and validity study. J Psychiatr Res 2018 Jan;96:239-246. [doi: 10.1016/j.jpsychires.2017.10.017] [Medline: 29126059]
34. Cella M, He Z, Killikelly C, Okruszek Ł, Lewis S, Wykes T. Blending active and passive digital technology methods to improve symptom monitoring in early psychosis. Early Interv Psychiatry 2019 Oct;13(5):1271-1275. [doi: 10.1111/eip.12796] [Medline: 30821079]
35. Schlosser D, Campellone TR, Truong B, Etter K, Vergani S, Komaiko K, et al. Efficacy of PRIME, a Mobile App Intervention Designed to Improve Motivation in Young People With Schizophrenia. Schizophr Bull 2018 Aug 20;44(5):1010-1020 [FREE Full text] [doi: 10.1093/schbul/sby078] [Medline: 29939367]
36. Bucci S, Barrowclough C, Ainsworth J, Machin M, Morris R, Berry K, et al. Actissist: Proof-of-Concept Trial of a Theory-Driven Digital Intervention for Psychosis. Schizophr Bull 2018 Aug 20;44(5):1070-1080 [FREE Full text] [doi: 10.1093/schbul/sby032] [Medline: 29566206]
37. Bucci S, Berry N, Morris R, Berry K, Haddock G, Lewis S, et al. "They Are Not Hard-to-Reach Clients. We Have Just Got Hard-to-Reach Services." Staff Views of Digital Health Tools in Specialist Mental Health Services. Front Psychiatry 2019;10:344 [FREE Full text] [doi: 10.3389/fpsyt.2019.00344] [Medline: 31133906]

Abbreviations

ACT-DL: acceptance and commitment therapy in daily life
CBT: cognitive behavioral therapy
CHR: clinical high risk

FEP: first-episode psychosis
PICO: P - problem/patient/population, I - intervention/indicator, C - comparison, O - outcome
TAU: treatment as usual

Edited by G Eysenbach; submitted 24.09.19; peer-reviewed by L Sheets, E Moran; comments to author 13.10.19; revised version received 23.10.19; accepted 26.10.19; published 12.11.19.

Please cite as:
Camacho E, Levin L, Torous J
Smartphone Apps to Support Coordinated Specialty Care for Prodromal and Early Course Schizophrenia Disorders: Systematic Review
J Med Internet Res 2019;21(11):e16393
URL: http://www.jmir.org/2019/11/e16393/
doi:10.2196/16393
PMID:31714250

Viewpoint

The Real Era of the Art of Medicine Begins with Artificial Intelligence

Bertalan Meskó[1], MD, PhD

The Medical Futurist Institute, Budapest, Hungary

Corresponding Author:
Bertalan Meskó, MD, PhD
The Medical Futurist Institute
Povl Bang-Jensen u. 2/B1. 4/1
Budapest, 1118
Hungary
Phone: 36 703807260
Email: berci@medicalfuturist.com

Abstract

Physicians have been performing the art of medicine for hundreds of years, and since the ancient era, patients have turned to physicians for help, advice, and cures. When the fathers of medicine started writing down their experience, knowledge, and observations, treating medical conditions became a structured process, with textbooks and professors sharing their methods over generations. After evidence-based medicine was established as the new form of medical science, the art and science of medicine had to be connected. As a result, by the end of the 20th century, health care had become highly dependent on technology. From electronic medical records, telemedicine, three-dimensional printing, algorithms, and sensors, technology has started to influence medical decisions and the lives of patients. While digital health technologies might be considered a threat to the art of medicine, I argue that advanced technologies, such as artificial intelligence, will initiate the real era of the art of medicine. Through the use of reinforcement learning, artificial intelligence could become the stethoscope of the 21st century. If we embrace these tools, the real art of medicine will begin now with the era of artificial intelligence.

(J Med Internet Res 2019;21(11):e16295) doi:10.2196/16295

KEYWORDS

future; artificial intelligence; digital health; technology; art of medicine

From the Dawn of Medicine to the 21st Century

Since Hippocrates, physicians have been performing the art of medicine. With limited knowledge, experience, and rudimentary tools, they have been the guardians of health and disease. The medical profession is highly regarded and has always been considered exceptional, and since the ancient era, patients have turned to physicians for help, advice, and cures. When the fathers of medicine started writing down their experience, knowledge, and observations, treating medical conditions became a structured process with textbooks and professors sharing their methods over generations.

In his essay “Teacher and Student”, Sir William Osler noted that: “The practice of medicine is an art, based on science,” [1]. Physicians had to possess the necessary skills to make use of their vast knowledge and experience, but they also had to have a spiritual understanding of treating the patient and not the disease [2]. Some argue that it is the mission of medical professionals to provide the art and science of medicine [3]. Then, at the dawn of modern medicine in the late 19th and early 20th century, there were attempts to transform the art of medicine into science to supplement clinical judgment using the conclusions of a scheme of logic [4]. After evidence-based medicine was established as the new form of medical science, it seemed that the art and science of medicine had to be connected [5]. Physicians were expected to treat the patient and not the disease, had to deal with a myriad of data about their patients and had to handle the increasing availability of new clinical science and new studies.

In the case of the first stethoscope, a hollow wooden tube introduced by Dr. Laennec in France in the early 19th century to augment cardiac and lung sounds, it took decades to spread the idea of improving care with it [6]. That one innovation needed so much time to reach the masses and become a common medical practice as physicians were reluctant to use it. Now dozens of innovations come out daily, and the burden and pressure are enormous on physicians to adopt them properly.

By the mid-20th century, health care had become highly dependent on technology. Physicians had to learn to work closely with it, which is consequentially a major contributor to their burnout today [7]. When personal computers became widely available, the concept of electronic health emerged. The digitalization of medical records became inevitable, as the amount of available medical knowledge continued to grow rapidly by millions of new studies every year, and patients started to become empowered to oversee their own health care. When personal computers could be connected to networks, telemedical services appeared. The rise of social media networks gave space for the rise of Medicine 2.0 and Health 2.0, while the penetration of mobile phones and later smartphones resulted in mobile health. However, since the 2010s, the rate at which disruptive technologies appear has become overwhelming for both patients and their caregivers [8]. This is the cultural transformation that, under the term digital health, has been shaping the fundamental structures of medicine and the doctor-patient relationship.

The doctor-patient hierarchy has been transforming into a partnership, with many formerly passive patients becoming proactive in their care and wanting to be involved in decision making. The ivory tower of medicine has been breaking down and opening for nonprofessionals too. Thus, digital health technologies could be considered a threat to the art of medicine [9], but I argue that it is not the case. On the contrary, advanced technologies and proactive patients will initiate the real era of the art of medicine.

Artificial Intelligence Will Facilitate the Rise of the Art of Medicine

Despite common fears about the perceived dangers of automation, especially artificial intelligence, it does not seem to be taking the jobs of physicians or monopolizing medicine. As studies show, it will instead help automate administrative tasks and take over monotonous day-to-day assignments [10]. It has the potential to free up time for medical professionals to let them fulfill the mission they signed up for: to help people on their health care journeys with compassion, creativity, and care.

The art of medicine requires attention, time, and empathy from physicians while they are treating patients [11]. With the challenging interfaces of electronic medical records, a lot of administrative tasks that need to be handled, and a constant pressure from payers impacting physicians, even the chance for the art of medicine vanishes. However, the use of artificial intelligence (AI) could help facilitate the art of medicine [12].

Advanced technologies such as AI will transform the role of physicians [13]. Medical professionals might prefer to be the translator of the technical data for the patient, act as a guide in the jungle of digital health for their patient, and be a counselor in navigating through health care choices, instead of being the ultimate source of medical knowledge and the sole decision-maker regarding medication and treatment choice. There are case studies that underscore this notion [14]. There was a game between the reigning chess world champion Garry Kasparov and IBM's Deep Blue supercomputer in 1997 that ended in a triumph for the algorithms/the programmers of the algorithm. It was believed then that this win might indicate the end of chess, because why would anyone play chess when the best player in the world was a program? However, over 600,000 million people play chess today, which is higher than in previous years. No professional player can prepare without working with computers now, as technology has made it more popular and increased the level at which players can play.

A similar path awaited the Chinese game, Go, which is one of the most complex games ever invented. In chess, the number of moves at any given moment in a game is approximately thirty. On the 19×19 square grid of the Go board, players can have around 200 moves, with the number of configurations on the board more than the number of atoms in the universe. This is only one of the factors that made the victory of Google Deepmind's AlphaGo over the 18-time Go World Champion, Lee Sedol, a milestone in computation and a spectacle for 80 million viewers worldwide [15]. Winning in Go requires creativity and intuition, human skills that are beyond current artificial intelligence technology. The key to success for AlphaGo was in the way in which it was trained to play the game [16], as it used a combination of neural networks and reinforcement learning. In reinforcement learning, developers don't tell the algorithm how to play the game and instead let it play millions of games against itself, only telling the algorithm which outcomes were the desired ones (the matches the algorithm could win) and letting the program find its strategy and rules for how to achieve those desired outcomes. This means that the cognitive limitations of people are not programmed into the algorithm. In theory, it could find new ways of playing the game even though it has been played by millions of people for thousands of years. Thus, such algorithms are not affected by the concepts, philosophies, beliefs, and mechanisms of human thinking, nor are they affected by the boundaries and flaws of human intelligence.

Advanced algorithms have already been shown to improve several aspects of practicing medicine [17], including radiology, oncology, cardiology, medical errors, and cost-effectiveness, among others. An algorithm could use reinforcement learning to find a treatment or even a cure for a medical condition that physicians and researchers cannot, and while the treatment or cure works, the underlying reasons why are not clear. Patients get better, get cured, or can manage their condition properly, but the way it was found is not clear or understood by those who administer the therapy. This would be the era of the modern art of medicine. Medical professionals using advanced technologies can care for their patients better, and their challenge is not in finding out how to care for their patients but instead to understand why certain methods work and others do not. When clinical science is not lagging behind the practice of medicine but progressing ahead of it, the art of medicine is going to be understanding how algorithms got to a conclusion.

Another way that AI could help physicians practice the art of medicine is if watched over these algorithms and evaluated outputs against their best clinical judgment as a safety check. Systems will inevitably get the answer wrong occasionally, and

we will need good clinical judgment to intervene when that occurs.

Preparing for the Modern Art of Medicine

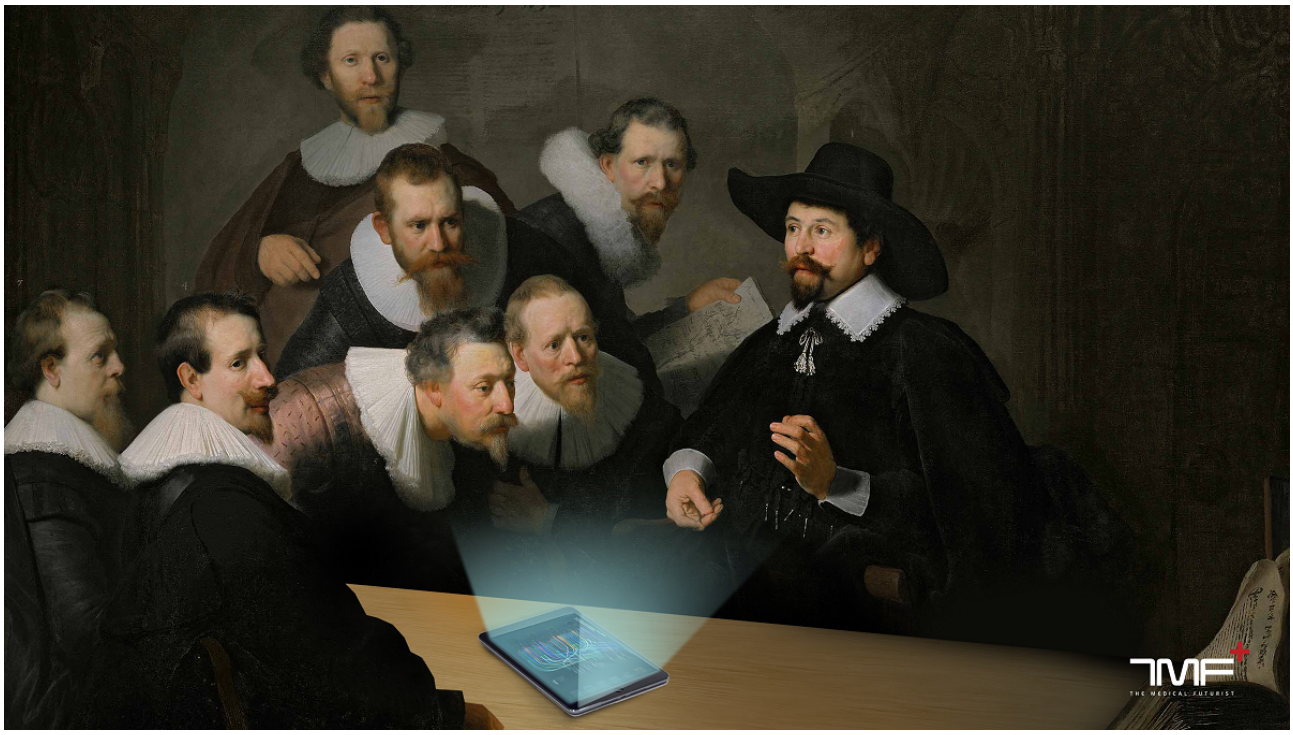
AlphaGo was able to beat the world champion four to one. Lee Sedol was deeply disappointed and felt embarrassed at first, as he felt he was representing humans in a losing battle against a technological entity. However, by the fifth game they played, it seemed that Sedol had a newfound understanding of the way AlphaGo evaluated the game and its positions. He stated that by doing that he became a much better player. He was able to stand on the shoulders of technology to further develop his skills and vision about the future of the game. This is exactly how the medical community should look at the impact digital health technologies have on health care and on their jobs.

In the next decade, it is believed that most patients will become empowered and proactive in their care, using sensors to obtain data, and using the information they find online as well as algorithms to understand this health data. Seamless interfaces and invisible technologies, such as chatbots with voice-to-text functions and algorithms scanning the medical literature, will become a common element of practicing medicine. Moreover, artificial narrow intelligence algorithms will discover new treatments and run *in silico* clinical trials that physicians, pharma companies, or medical innovators would never think of. As it will not be limited by the traditional pathways and thought patterns used for centuries in medicine, artificial intelligence may produce entirely new solutions for tackling global health issues.

The supercomputer in Douglas Adam's "The Hitchhiker's Guide To The Galaxy", after thinking for seven and a half million years, provided the answer to life, to the universe, and everything: 42 [18]. What if smart algorithms just started spitting out the answers to big questions without any explanation? The real art of medicine would thus be figuring out the logical path of how an AI arrived at a certain solution. That will require high levels of creativity, problem-solving, and cognitive skills from the medical community.

I believe that AI is going to be the stethoscope of the 21st century. While digital health technologies provide us with more health data than ever before and AI helps analyze it to improve health and well-being, to cut down on administrative tasks, or even to optimize both physicians' and patients' schedules, we should never forget that they are going to be tools in the hands of physicians and not the other way around [19]. We can be confident that compassionate care, empathy, creativity, problem-solving, and profound human connection will never cease to be at the core of caring for patients. If we embrace these tools through well-designed policies, constant education, and proper guidelines, the real art of medicine will begin now with the era of artificial intelligence (Figure 1).

Figure 1. A concept art about the art of medicine in the era of artificial intelligence based on Rembrandt's painting The Anatomy Lesson of Dr Nicolaes Tulp.



Conflicts of Interest

None declared.

References

1. Van Der Weyden M. The essence of the art of medicine. Medical Journal of Australia 2006 Aug 21;185(4):185-185 [FREE Full text] [doi: 10.5694/j.1326-5377.2006.tb00525.x]
2. Mambu J. Restoring the Art of Medicine. Am J Med 2017 Dec;130(12):1340-1341. [doi: 10.1016/j.amjmed.2017.06.034] [Medline: 28757320]
3. Nelson AR. Humanism and the Art of Medicine. JAMA 1989 Sep 01;262(9):1228. [doi: 10.1001/jama.1989.03430090090040]
4. Komaroff AL. Algorithms and the "art" of medicine. Am J Public Health 1982 Jan;72(1):10-12. [doi: 10.2105/ajph.72.1.10] [Medline: 7053613]
5. Parker M. False dichotomies: EBM, clinical freedom, and the art of medicine. Med Humanit 2005 Jun 15;31(1):23-30 [FREE Full text] [doi: 10.1136/jmh.2004.000195] [Medline: 16167410]
6. Cheng TO. How Laënnec invented the stethoscope. Int J Cardiol 2007 Jun 12;118(3):281-285. [doi: 10.1016/j.ijcard.2006.06.067] [Medline: 17046081]
7. Shanafelt TD, Dyrbye LN, Sinsky C, Hasan O, Satele D, Sloan J, et al. Relationship Between Clerical Burden and Characteristics of the Electronic Environment With Physician Burnout and Professional Satisfaction. Mayo Clin Proc 2016 Jul;91(7):836-848. [doi: 10.1016/j.mayocp.2016.05.007] [Medline: 27313121]
8. Meskó B, Drobni Z, Bényei É, Gergely B, Gyorffy Z. Digital health is a cultural transformation of traditional healthcare. Mhealth 2017;3:38 [FREE Full text] [doi: 10.21037/mhealth.2017.08.07] [Medline: 29184890]
9. Eledrisi MS. We must save the art of medicine. Arch Intern Med 2000 Jun 12;160(11):1701-1702. [doi: 10.1001/archinte.160.11.1701] [Medline: 10847267]
10. Vegard KR, Amico R, Thomas RJ. Accenture Institute for Higher Performance. 2016. The Promise of Artificial Intelligence URL: https://www.accenture.com/t20170411t174532z__w__/mu-en/_acnmedia/pdf-32/ai_in_management_report.pdf [accessed 2019-11-12]
11. Meskó B, Radó N, Gyorffy Z. Opinion leader empowered patients about the era of digital health: a qualitative study. BMJ Open 2019 Mar 20;9(3):e025267 [FREE Full text] [doi: 10.1136/bmjopen-2018-025267] [Medline: 30898816]
12. Shen J, Zhang CJP, Jiang B, Chen J, Song J, Liu Z, et al. Artificial Intelligence Versus Clinicians in Disease Diagnosis: Systematic Review. JMIR Med Inform 2019 Aug 16;7(3):e10010 [FREE Full text] [doi: 10.2196/10010] [Medline: 31420959]
13. Meskó B, Gyorffy Z. The Rise of the Empowered Physician in the Digital Health Era: Viewpoint. J Med Internet Res 2019 Mar 26;21(3):e12490 [FREE Full text] [doi: 10.2196/12490] [Medline: 30912758]
14. Meskó B, Radó N. The Medical Futurist. 2019 Apr 30. The Real Era of The Art of Medicine Begins With Artificial Intelligence URL: https://medicalfuturist.com/artificial-intelligence-and-the-art-of-medicine/ [accessed 2019-11-12]
15. Kohs G. AlphaGo [DVD]. 2017. URL: https://www.alphagomovie.com/ [accessed 2019-11-12]
16. Sadler M, Regan N. Game Changer: Alphazero's Groundbreaking Chess Strategies And The Promise Of AI. Alkmaar, Netherlands: New In Chess; 2019.
17. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med 2019 Jan 7;25(1):44-56. [doi: 10.1038/s41591-018-0300-7] [Medline: 30617339]
18. Adams D. The Hitchhiker's Guide To The Galaxy. New York City, New York: Pan Macmillan; 2009.
19. Oh S, Kim JH, Choi S, Lee HJ, Hong J, Kwon SH. Physician Confidence in Artificial Intelligence: An Online Mobile Survey. J Med Internet Res 2019 Mar 25;21(3):e12422 [FREE Full text] [doi: 10.2196/12422] [Medline: 30907742]

Abbreviations

AI: artificial intelligence

Edited by G Eysenbach; submitted 17.09.19; peer-reviewed by L Buis; comments to author 18.10.19; revised version received 04.11.19; accepted 11.11.19; published 18.11.19.

Please cite as:
Meskó B
The Real Era of the Art of Medicine Begins with Artificial Intelligence
J Med Internet Res 2019;21(11):e16295
URL: http://www.jmir.org/2019/11/e16295/
doi:10.2196/16295
PMID:31738169

Original Paper

Assessing the Psychometric Properties of the Digital Behavior Change Intervention Engagement Scale in Users of an App for Reducing Alcohol Consumption: Evaluation Study

Olga Perski[1], PhD; Jim Lumsden[2], PhD; Claire Garnett[1], PhD; Ann Blandford[3], PhD; Robert West[1], PhD; Susan Michie[4], DPhil

[1]Department of Behavioural Science and Health, University College London, London, United Kingdom

[2]UK Centre for Tobacco and Alcohol Studies, School of Experimental Psychology, University of Bristol, Bristol, United Kingdom

[3]UCL Interaction Centre, University College London, London, United Kingdom

[4]Department of Clinical, Educational and Health Psychology, University College London, London, United Kingdom

Corresponding Author:
Olga Perski, PhD
Department of Behavioural Science and Health
University College London
1-19 Torrington Place
London, WC1E 7HB
United Kingdom
Phone: 44 20 7679 1258
Email: olga.perski@ucl.ac.uk

Abstract

Background: The level and type of engagement with digital behavior change interventions (DBCIs) are likely to influence their effectiveness, but validated self-report measures of engagement are lacking. The DBCI Engagement Scale was designed to assess behavioral (ie, amount, depth of use) and experiential (ie, attention, interest, enjoyment) dimensions of engagement.

Objective: We aimed to assess the psychometric properties of the DBCI Engagement Scale in users of a smartphone app for reducing alcohol consumption.

Methods: Participants (N=147) were UK-based, adult, excessive drinkers recruited via an online research platform. Participants downloaded the *Drink Less* app and completed the scale immediately after their first login in exchange for a financial reward. Criterion variables included the objectively recorded amount of use, depth of use, and subsequent login. Five types of validity (ie, construct, criterion, predictive, incremental, divergent) were examined in exploratory factor, correlational, and regression analyses. The Cronbach alpha was calculated to assess the scale's internal reliability. Covariates included motivation to reduce alcohol consumption.

Results: Responses on the DBCI Engagement Scale could be characterized in terms of two largely independent subscales related to experience and behavior. The experiential and behavioral subscales showed high (α=.78) and moderate (α=.45) internal reliability, respectively. Total scale scores predicted future behavioral engagement (ie, subsequent login) with and without adjusting for users' motivation to reduce alcohol consumption (adjusted odds ratio [OR_{adj}]=1.14; 95% CI 1.03-1.27; *P*=.01), which was driven by the experiential (OR_{adj}=1.19; 95% CI 1.05-1.34; *P*=.006) but not the behavioral subscale.

Conclusions: The DBCI Engagement Scale assesses behavioral and experiential aspects of engagement. The behavioral subscale may not be a valid indicator of behavioral engagement. The experiential subscale can predict subsequent behavioral engagement with an app for reducing alcohol consumption. Further refinements and validation of the scale in larger samples and across different DBCIs are needed.

(J Med Internet Res 2019;21(11):e16197) doi:10.2196/16197

KEYWORDS

engagement; digital behaviour change interventions; mHealth; psychometrics; self-report scale; smartphone apps; excessive alcohol consumption

Introduction

Some level of engagement with digital behavior change interventions (DBCIs) is necessary for the effectiveness of such interventions [1]. However, observed levels of engagement with DBCIs are often considered too limited to support behavior change [2]. For example, a systematic review of Web-based health interventions found that approximately 50% of participants engaged with the interventions in the desired manner, with estimates varying between 10% and 90% across trials [3]. Studies conducted across different settings and target behaviors report a positive association of DBCI engagement and intervention effectiveness [4,5], suggesting that these variables may be linked via a dose-response function [1,6]. However, it is also plausible that individuals who are more successful in achieving change in the behavior targeted by the DBCI engage with DBCIs more [7] or that a limited amount of engagement is sufficient for bringing about meaningful change in some users (ie, "effective engagement") [6]. Attempts have been made to characterize the function linking engagement with intervention effectiveness [1,7-9], but progress is hindered due to the use of different definitions and measures of engagement across studies.

The question of what it means for someone to be engaged with a DBCI has been of interest to psychologists and computer scientists alike. Broadly, psychologists have defined engagement as the extent of technology use, perceived as a proxy for participant exposure to a DBCI's "active ingredients" or component behavior change techniques [10,11]. On the other hand, computer scientists have defined engagement as the subjective experience of "flow" or "immersion" that occurs during the human-computer interaction, characterized by focused attention, intrinsic interest, balance between challenge and skill, losing track of time and self-consciousness, and transportation to a "different place" [12,13]. After having conducted a systematic, integrative literature review of the psychology and computer science literatures [7] in addition to in-depth interviews with potential DBCI users, our interdisciplinary research team proposed the following working definition of engagement: "[Engagement with a DBCI is] a state-like construct which occurs each time a user interacts with a DBCI, with two behavioral (ie, amount and depth of use) and three experiential (ie, attention, interest and enjoyment) dimensions [14]."

We hence theorized that two behavioral (amount and depth of use) and three experiential (attention, interest, and enjoyment) dimensions are necessary and sufficient conditions for someone to be engaged with a DBCI. Although similar, engagement with DBCIs is thought to be conceptually distinct from both "flow" and pure technology usage. Although several measures of flow, immersion, and technology usage are available for use (for overviews, see [7,14,15]), an instrument that quantifies the intensity of behavioral and experiential engagement is lacking. For a quantitative scale of engagement to be useful for researchers, practitioners, and developers, it should be able to predict key variables of interest such as future engagement, knowledge acquisition, or intervention effectiveness. In addition, although a number of usage metrics derived from log-data are typically used to capture the intensity of behavioral engagement [15-17], a validated measure of engagement, which captures both the experiential and behavioral dimensions of engagement and could be easily administered without the need to access and process the DBCI's raw data, would be useful. The DBCI Engagement Scale was developed to fill this gap [14].

As part of the scale development process (described in detail in [14]), a pool of initial scale items was developed by the interdisciplinary research team in addition to two "best bets" for a short measure of engagement. Lay and expert respondents were then asked to classify the initial scale items into one of six categories (ie, *amount of use*, *depth of use*, *interest*, *attention*, *enjoyment*, plus an *unclassified* category) to examine the scale's content validity. The first psychometric evaluation of the 10-item DBCI Engagement Scale was conducted in a sample of adult excessive drinkers who had voluntarily downloaded a freely available, evidence-informed app—*Drink Less*—for reducing their alcohol consumption [14]. Results indicated that the behavioral and experiential indicators of engagement may resolve to a single dimension. However, fewer than 5% of eligible users completed the scale during the study, and a sensitivity analysis indicated that the analytic sample was biased toward highly engaged users.

Studying engagement in real-world settings is notoriously difficult, as highly engaged users are more likely to respond to research surveys [18], potentially biasing results. Moreover, evidence suggests that motivation to change the target behavior is consistently associated with the frequency of behavioral engagement, such as the total number of logins [19,20]. Although motivation to change is a key predictor of engagement, it is neither a necessary nor a sufficient condition for someone to be engaged with a DBCI. For example, a user with low motivation to reduce their alcohol consumption might be intrigued by the design of a specific app, engage with its content, and subsequently become motivated to drink less. Therefore, to better study the dimensional structure of engagement, we considered it important to adjust for motivation to change in our analyses, thus separating the state of engagement from confounding motivations. This study aimed to evaluate the DBCI Engagement Scale in a sample of users recruited via an online research platform in order to address the following research questions:

1. What is the factor structure of the DBCI Engagement Scale? (construct validity)
2. Is the DBCI Engagement Scale internally reliable? (internal reliability)
3. Are total scale scores positively associated with objectively recorded amount of use and depth of use? (criterion validity)
4. Do total scale scores predict future behavioral engagement (ie, subsequent login), with and without adjustment for motivation to reduce alcohol consumption? (predictive validity)
5. Do two best bets for a short measure of engagement predict future behavioral engagement, with and without adjustment for motivation to reduce alcohol consumption? (predictive validity)
6. Does a model including the objectively recorded behavioral and the self-reported experiential indicators of engagement

account for more variance in future behavioral engagement (ie, subsequent login) compared with a model including only the objectively recorded behavioral indicators of engagement? (incremental validity)

7. Are total scale scores significantly associated with scores on the Flow State Scale? (divergent validity)

Methods

The preregistered study protocol can be found in the Open Science Framework [21]. Ethical approval was granted by University College London's Computer Science Departmental Research Ethics Chair (Project ID: UCLIC/1617/004/Staff Blandford HFDH).

Inclusion Criteria

Participants were eligible to take part in the study if they were aged ≥18 years; reported an Alcohol Use Disorders Identification Test (AUDIT) score ≥8, indicating excessive alcohol consumption [22]; were residing in the United Kingdom; owned an iPhone capable of running iOS 8.0 software (ie, an iPhone 4S or later models); and were willing to download and explore an app for reducing alcohol consumption.

Sampling

Participants were recruited via the online research platform Prolific [23]. Individuals who take part in research via online platforms are primarily motivated by the financial incentives and are not necessarily interested in health behavior change [24]. Therefore, we expected that Prolific would enable us to recruit a sample of users with different levels of motivation to change. We did not, however, expect to recruit a sample that is representative of the general population of excessive drinkers in the United Kingdom.

Sample Size

No formal sample size calculation was performed. Based on the psychometric literature, a 25:1 participant-to-item ratio (ie, a total of 250 participants) was considered desirable [25].

Measures

To determine eligibility and describe the sample, data were collected on age; sex (female or male); type of work (manual, nonmanual, or other); patterns of alcohol consumption, measured by the AUDIT [22]; motivation to reduce alcohol consumption, measured by the Motivation To Stop Scale [26-28]; country of residence (United Kingdom or other); iPhone ownership (yes or no); and willingness to download and explore an alcohol reduction app (yes or no).

For eligible participants who downloaded and explored the *Drink Less* app, data were collected on location during first use of the app (home, work, vehicle, public transport, restaurant/pub/café, other's home, other, or can't remember) and the 10-item DBCI Engagement Scale [14], which captures momentary behavioral (ie, amount, depth of use) and experiential (ie, attention, interest, enjoyment) engagement with DBCIs (Textbox 1). A detailed account of how the scale items were developed and tested in a group of experts and nonexperts can be found in a previous study [14].

Data were also collected on the below variables, which were used to test the scale's criterion, predictive, incremental, and divergent validity.

Textbox 1. The DBCI Engagement Scale.

Please answer the following questions with regard to your most recent use of the *Drink Less* app.

How strongly did you experience the following?

1. Interest
2. Intrigue
3. Focus
4. Inattention
5. Distraction
6. Enjoyment
7. Annoyance
8. Pleasure *(Measured on a 7-point scale with end- and middle-points anchored: "not at all," "moderately," and "extremely")*
9. How much time (in minutes) do you roughly think that you spent on the app? *(Enter free text)*
10. Which of the app's components do you remember visiting? (You can select multiple options)

a) Calendar *(coded by the researchers as "Self-Monitoring/Feedback")*

b) Create and view goals *(coded as "Goal Setting")*

c) What has and hasn't worked *(coded as "Self-Monitoring/Feedback")*

d) Create and view action plans *(coded as "Action Planning")*

e) Your hangover and you *(coded as "Self-Monitoring/Feedback")*

f) Review your drinking *(coded as "Normative Feedback")*

g) Dashboard *(coded as "Self-Monitoring/Feedback")*

h) Game *(coded as "Cognitive Bias Re-Training")*

i) Drink + me *(coded as "Identity Change")*

j) Useful information *(coded as "Other")*

k) Other *(coded as "Other")*

l) Can't remember *(coded as "Other")*

Indexed as a proportion of available modules (eg, 5/7×100=71.4).

Construct, Criterion, and Incremental Validity

A record of the number of app screens viewed was kept during participants' first login session to derive the objectively recorded *amount of use* and *depth of use*, which were used to test the scale's construct, criterion, and incremental validity. The screen view records were stored in an online database (*NodeChef*) and extracted using the free python library *pandas*. The variable *amount of use* was derived by calculating the time spent (in seconds) during participants' first login session. The variable *depth of use* was derived by calculating the number of app components visited during participants' first login session, indexed as a proportion (0-100) of the number of available components within the *Drink Less* app (ie, Goal Setting, Self-monitoring/Feedback, Action Planning, Normative Feedback, Cognitive Bias Re-Training, Identity Change, Other [29]).

Predictive Validity

A record of the number of app screens viewed was also kept over the next 14 days to derive the variable *subsequent login*, which was used to test the scale's predictive validity. A subsequent login (yes vs no) was defined as a new screen view following at least 30 minutes of inactivity [30]. As health apps are likely to be abandoned after users' first login [31,32], the authors theorized that a useful measure of engagement should be able to distinguish between users who are likely to return to an app.

Two items that represented the authors' *best bets* for a short measure of engagement (ie, "How much did you like the app?" and "How engaging was the app?") were developed by the study team and used to test whether a short measure of engagement had superior predictive validity compared with the scale in its entirety. These items were not explicitly drawn from published self-report scales.

Divergent Validity

Two items from the Flow State Scale [33] were used to test the scale's divergent validity. We selected two items that were previously found to load most strongly onto the general *flow* factor (ie, "When using *Drink Less*, the way time passed seemed to be different from normal," "When using *Drink Less*, I was not worried about what others may have been thinking of me"). Although there is some overlap in the experiential indicators of the states of engagement and flow (ie, focused attention, interest), the study team theorized that users do not necessarily

experience *loss of time and consciousness* or *balance between challenge and skill* when engaging with a DBCI. Assessing whether users can be engaged without necessarily being in a state of flow was therefore considered a useful test of the scale's divergent validity. The Flow State Scale has previously been applied in the context of digital gaming [34].

Procedure

Interested participants were identified via the recruitment platform, Prolific, and received a compensation of £0.50 for completing the screening questionnaire, hosted by Qualtrics survey software (Provo, Utah). Eligible participants were invited via Prolific's internal email system and asked to download the *Drink Less* app from the Apple App Store. Participants were instructed to explore the *Drink Less* app in the way that they would explore any new app and were told that the researchers would monitor their app usage to assess what content they were interested in. For technical reasons, participants were told that they had to select the option *Interested in drinking less alcohol* when asked about why they were using the *Drink Less* app and to enable the push notifications. When clicking on the phone's home button after having finished exploring the app, participants received a push notification with a link to the study survey. Participants were subsequently asked to enter their Prolific identification number, which enabled the researchers to match participants' survey responses to their app screen views. Participants who initiated but did not complete the study survey (as indicated by their response status on Prolific's platform, which was either labelled "Timed out" or "Returned submission") were sent one reminder message. On completing the task, participants were paid £1.25.

Data Analysis

All analyses were conducted in SPSS version 20.0 (IBM Corporation, Armonk, New York). The assumptions for parametric tests were assessed (ie, normality of the distribution of residuals) and when violated, normalization was applied (ie, *z*-score normalization of positively skewed data). Descriptive statistics (eg, mean, range, variance) were calculated for each scale item and the criterion variables of interest to determine suitability for factor analysis.

Construct Validity

It was hypothesized that a five-factor solution (ie, *amount of use*, *depth of use*, *attention*, *interest*, *enjoyment*) would provide the best fit of the observed data [14]. A series of exploratory factor analyses (EFAs) using principal axis factoring estimation and oblique rotation was conducted. The inspection of Cattell's scree plots and the Kaiser criterion (ie, factors with eigenvalues >1) was used to determine the number of factors to retain [25]. First, we tested the fit of a solution including the self-reported items. This was compared with a solution including a combination of the self-reported indicators of experiential engagement and the objectively recorded indicators of behavioral engagement (ie, objective *amount of use* and *depth of use*).

Internal Reliability

Internal consistency reliability was assessed by calculating the Cronbach alpha. A large coefficient (ie, =.70 or above) was interpreted as evidence of strong item covariance [35].

Criterion Validity

Criterion validity was assessed by calculating the Pearson correlation coefficient for the relationship between participants' automatically recorded app screen views from their first login (ie, objective *amount of use* and *depth of use*) with their self-reported *amount of use* and *depth of use* and their total scale scores.

Predictive Validity

The variable *subsequent login* was regressed onto participants' total scale scores, with and without adjustment for motivation to reduce alcohol consumption.

The variable *subsequent login* was also regressed onto each of the two *best bets* for a short measure of engagement (ie, "How engaging was the app?" and "How much did you like the app?"), with and without adjustment for motivation to reduce alcohol consumption.

Incremental Validity

Incremental validity was assessed in two steps. First, we assessed the variance accounted for in the variable *subsequent login* by the objectively recorded indicators of behavioral engagement. This was compared with the variance accounted for in the variable *subsequent login* after adding the self-reported indicators of experiential engagement to the objectively recorded indicators of behavioral engagement.

Divergent Validity

Divergent validity was assessed by calculating the Pearson correlation coefficient for the relationship between each of the two indicators of the state of flow from the Flow State Scale [33] and the overall measure of engagement.

Results

Participants

During the study period (31 days; July 23, 2018 to August 22, 2018), 401 participants completed the online screening survey, of which 266 were eligible to take part. Of these, 147 (55%) participants downloaded the *Drink Less* app and completed the task (Figure 1). Due to funding restrictions, we were unable to extend the recruitment beyond this time point. The desired target sample size of 250 participants was hence not achieved. Participants' demographic and drinking characteristics are reported in Table 1. We did not detect any significant differences between eligible participants who did and did not complete the task on the demographic characteristics assessed.

Figure 1. Participant flow chart. AUDIT: Alcohol Use Disorders Identification Test; ID: identification.

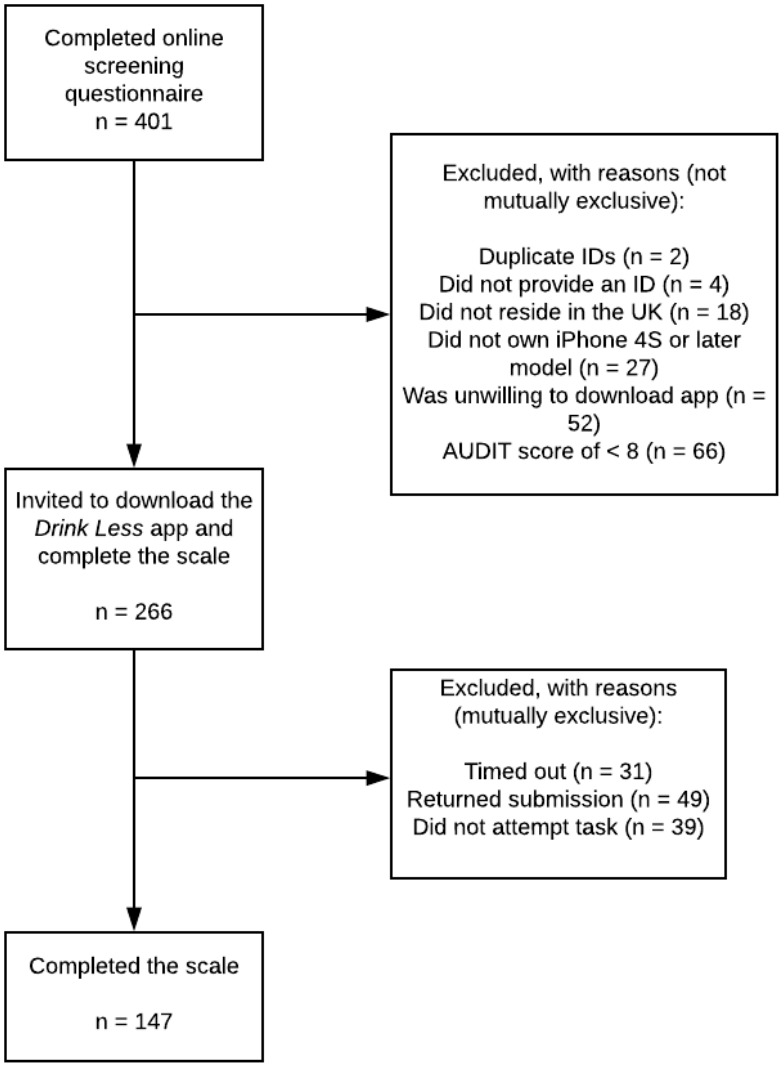


Table 1. Participants' demographic and drinking characteristics.

Demographic characteristics	Completed scale (n=147)	Eligible but did not complete scale (n=119)	*P* value[a]
Female gender, n (%)	97 (66)	71 (60)	.29
Type of work, n (%)			**.57**
Manual, n (%)	19 (13)	16 (13)	
Nonmanual, n (%)	89 (61)	78 (66)	
Other, n (%)	39 (27)	25 (21)	
Age (years), mean (SD)	34.4 (10.4)	36.6 (11.8)	.11
Drinking characteristics			
Motivation To Stop Scale, n (%)			**.08**
I don't want to cut down on drinking alcohol	14 (10)	26 (22)	
I think I should cut down on drinking alcohol but I don't really want to	43 (29)	25 (21)	
I want to cut down on drinking alcohol but I haven't thought about when	19 (13)	17 (14)	
I really want to cut down on drinking alcohol but I don't know when I will	17 (12)	11 (9)	
I want to cut down on drinking and hope to soon	23 (16)	17 (14)	
I really want to cut down on drinking alcohol and intend to in the next 3 months	11 (7)	4 (3)	
I really want to cut down on drinking alcohol and intend to in the next month	20 (14)	19 (16)	
Alcohol Use Disorders Identification Test, mean (SD)	15.4 (5.1)	14.2 (5.7)	.07

[a]Differences between groups were assessed using chi-square tests or *t* tests, as appropriate.

Descriptive Statistics

Descriptive statistics for the scale items are reported in Table 2. The majority of participants completed the scale at home (118/147, 80.3%) or work (19/147, 12.9%). To account for the observed skewness, z-score normalization was applied to the 10-scale items and the two items used for testing the scale's criterion validity. Inter-item correlations of the normalized scale items are reported in Table 3.

Scale Evaluation

Construct Validity

The Keiser-Meier Olkin Test of Sampling Adequacy (0.70) and the Bartlett Test of Sphericity ($P<.001$) indicated that data were suited for factor analysis. Three EFA solutions were tested to arrive at a best-fitting solution.

Table 2. Descriptive statistics for the scale items (N=147).

Statistic	Range	Mean (SD)	Variance	Skewness	Kurtosis
DBCI[a] Engagement Scale Items					
1. "How strongly did you experience interest?"	2-7	5.30 (1.09)	1.18	–0.30	0.06
2. "How strongly did you experience intrigue?"	1-7	5.39 (1.27)	1.61	–0.85	0.50
3. "How strongly did you experience focus?"	2-7	5.31 (1.18)	1.40	–0.56	0.14
4. "How strongly did you experience inattention?"[b]	1-7	5.61 (1.33)	1.76	–1.24	1.47
5. "How strongly did you experience distraction?"[b]	1-7	5.47 (1.45)	2.10	–1.12	0.86
6. "How strongly did you experience enjoyment?"	1-7	4.46 (1.44)	2.07	–0.10	–0.48
7. "How strongly did you experience pleasure?"	1-7	3.56 (1.64)	2.67	0.36	–0.70
8. "How strongly did you experience annoyance?"[b]	1-7	5.59 (1.39)	1.93	–1.09	1.08
9. "Which of the app's components did you visit?"	14.29-100.00	58.70 (22.00)	484.01	–0.12	–0.67
10. "How much time do you roughly think that you spent on the app?" (seconds)	120-1200	520.82 (237.21)	56,267.82	0.93	0.96
Variables used to test the scale's construct, criterion, and incremental validity					
11. Objectively recorded depth of use	28.57-100.00	66.66 (20.50)	420.28	–0.23	–0.85
12. Objectively recorded amount of use (seconds)	95.00-3,571.00	409.45 (360.71)	130,116.72	5.13	40.34
Items used to test the scale's divergent validity					
13 "When using *Drink Less*, the way time passed seemed different from normal."	1-5	2.76 (0.79)	0.62	0.11	0.10
14. "When using *Drink Less*, I was not worried about what others may have been thinking about me."	1-5	3.34 (1.16)	1.35	–0.24	–1.11
Variables/items used to test the scale's predictive validity					
15. "How much did you like the app?"	1-7	5.14 (1.29)	1.66	–0.80	0.82
16. "How engaging was the app?"	1-7	5.20 (1.17)	1.37	–0.65	0.66
17. Subsequent login (yes vs no), n (%)	67 (46)	N/A[c]	N/A	N/A	N/A

[a]DBCI: digital behavior change intervention.

[b]Values were reverse scored prior to the calculation of descriptive statistics.

[c]Not applicable.

Table 3. Inter-item correlation matrix (N=147).

DBCI Engagement Scale items	1[a]	2[b]	3[c]	4[d,e]	5[e,f]	6[g]	7[h]	8[e,i]	9[j]	10[k]	11[l,m]	12[m,n]
1. Interest	**1**											
P value	N/A[o]											
2. Intrigue	**0.44**	**1**										
P value	<.001	N/A										
3. Focus	**0.65**	**0.46**	**1**									
P value	<.001	<.001	N/A									
4. Inattention[e]	**0.18**	**0.10**	**0.31**	**1**								
P value	.027	.21	<.001	N/A								
5. Distraction[e]	**0.18**	**0.12**	**0.28**	**0.43**	**1**							
P value	.026	.16	.001	<.001	N/A							
6. Enjoyment	**0.48**	**0.31**	**0.44**	**–0.05**	**–0.15**	**1**						
P value	<.001	<.001	<.001	.59	.071	N/A						
7. Pleasure	**0.19**	**0.09**	**0.15**	**–0.19**	**–0.24**	**0.54**	**1**					
P value	.025	.30	.079	.019	.003	<.001	N/A					
8. Annoyance[e]	**0.28**	**0.15**	**0.37**	**0.27**	**0.24**	**0.29**	**0.12**	**1**				
P value	.001	.07	<.001	.001	.004	<.001	.16	N/A				
9. Which of app's components	**0.18**	**0.00**	**0.06**	**0.13**	**–0.03**	**0.19**	**0.19**	**.13**	**1**			
P value	.028	.97	.469	.11	.75	.019	.02	.13	N/A			
10. How much time spent	**0.10**	**0.10**	**–0.03**	**0.08**	**0.11**	**0.15**	**0.33**	**0.09**	**0.29**	**1**		
P value	.23	.22	.681	.33	.18	.07	<.001	.31	<.001	N/A		
11. Objective depth of use[m]	**0.13**	**0.11**	**0.15**	**0.18**	**0.01**	**0.11**	**–0.01**	**0.24**	**0.51**	**0.16**	**1**	
P value	.12	.18	.069	.03	.90	.20	.94	.003	<.001	.051	N/A	
12. Objective amount of use[m]	**0.31**	**0.18**	**0.28**	**0.16**	**0.06**	**0.25**	**0.00**	**0.19**	**0.10**	**0.10**	**0.52**	**1**
P value	<.001	.03	.001	.047	.49	.002	.97	.02	.22	.23	<.001	N/A

[a]Interest.
[b]Intrigue.
[c]Focus.
[d]Inattention.
[e]Values were reverse scored prior to analysis.
[f]Distraction.
[g]Enjoyment.
[h]Pleasure.
[i]Annoyance.
[j]Which of the app's components.
[k]How much time spent.
[l]Objective depth of use.
[m]Variables used to test the scale's construct, criterion, and incremental validity.
[n]Objective amount of use.
[o]Not applicable.

Solution 1

An EFA with oblique rotation was conducted. The eigenvalues indicated that a three-factor solution, accounting for 61.2% of the variance, was most appropriate (Table 4). The loadings indicated that the second factor comprised two of the negatively worded indicators (ie, items 4 and 5). The third factor comprised the two behavioral indicators (ie, items 9 and 10) and one of the experiential indicators (ie, item 7), which made little theoretical sense [14]. The loading of item 8 (also a negatively worded item) onto factor 1 was modest. Therefore, the negatively worded items (ie, items 4, 5, and 8) and item 7 were discarded prior to conducting a second EFA.

Solution 2

A subsequent EFA with oblique rotation indicated that a two-factor solution accounted for 62.4% of the variance (Table 4). The experiential indicators loaded clearly onto factor 1, and the behavioral indicators loaded clearly onto factor 2, with no cross-loadings (ie, items that load at 0.32 or higher on two or more factors) [25]. The two latent factors were labelled *Experiential Engagement* and *Behavioral Engagement*, respectively.

Solution 3

An EFA with oblique rotation using a combination of the self-reported experiential indicators (ie, items 1, 2, 3, and 6) and the automatically recorded behavioral indicators (ie, items 11 and 12) suggested a two-factor solution, which accounted for 65.7% of the variance. The experiential indicators loaded clearly onto factor 1, and the behavioral indicators loaded clearly onto factor 2 (Table 4).

Solution 2 was selected for use in the subsequent reliability and validity analyses, as it contained only the self-reported items and provided a similarly good fit of the data as Solution 3. A total scale score was calculated for each participant, with equal weight given to each of the retained items (ie, items 1, 2, 3, 6, 9, and 10).

Table 4. Factor loadings of the DBCI Engagement Scale in exploratory factor analyses.

Scale Items	Solution 1[a]			Solution 2[b]		Solution 3[c]	
	Factor 1	Factor 2	Factor 3	Factor 1	Factor 2	Factor 1	Factor 2
1. Interest	0.75[d]	0.14	0.25	0.80[d]	0.26	0.82[d]	0.28
2. Intrigue	0.51[d]	0.09	0.11	0.55[d]	0.09	0.55[d]	0.18
3. Focus	0.87[d]	0.28	0.09	0.83[d]	0.02	0.80[d]	0.27
4. Inattention[e]	0.25	0.61[d]	0.14	N/A[f]	N/A	N/A	N/A
5. Distraction[e]	0.21	0.68[d]	0.06	N/A	N/A	N/A	N/A
6. Enjoyment	0.66[d]	–0.35	0.43[d]	0.57[d]	0.31	0.57[d]	0.23
7. Pleasure	0.31	–0.48	0.56[d]	N/A	N/A	N/A	N/A
8. Annoyance[e]	0.41[d]	0.23	0.25	N/A	N/A	N/A	N/A
9. Which of app's components	0.16	0.01	0.43[d]	0.15	0.55	N/A	N/A
10. How much time spent	0.10	0.03	0.64[d]	0.09	0.53	N/A	N/A
11. Objective depth of use	N/A	N/A	N/A	N/A	N/A	0.37	0.77[d]
12. Objective amount of use	N/A	N/A	N/A	N/A	N/A	0.18	0.68[d]

[a]Exploratory factor analysis with oblique rotation, including items 1-10.

[b]Exploratory factor analysis with oblique rotation, including items 1, 2, 3, 6, 9, and 10.

[c]Exploratory factor analysis with oblique rotation, including items 1, 2, 3, 6, 11, and 12.

[d]Values with factor loadings ≥0.40.

[e]Values were reverse scored prior to analysis.

[f]Not applicable.

Internal Reliability

The internal consistency of the overall measure was 0.67, indicating moderate internal reliability [35]. The *Experiential Engagement* subscale had an internal consistency of 0.78, while the *Behavioral Engagement* subscale had an internal consistency of 0.45. Both subscales were significantly correlated with the measure overall (r_{145}=0.90, P<.001 and r_{145}=0.56, P<.001, respectively). However, the subscales were not significantly correlated with each other (r_{145}=0.15, P=.07).

Criterion Validity

Total scale scores were significantly correlated with objectively recorded *depth of use* (r_{145}=0.32, P<.001) and objectively recorded *amount of use* (r_{145}=0.33, P<.001). Self-reported *depth of use* was significantly correlated with objectively recorded

depth of use (r_{145}=0.51, *P*<.001). Self-reported *amount of use* was not significantly correlated with objectively recorded *amount of use* (r_{145}=0.10, *P*=.23).

Predictive Validity

Results from the predictive validity analyses are presented in Table 5. In the unadjusted analysis, total scale scores were significantly associated with future behavioral engagement, ie, the variable *subsequent login* (odds ratio [OR]=1.15, 95% CI 1.05-1.27, *P*=.01). The association remained significant in the model adjusting for motivation to reduce alcohol consumption (adjusted OR [OR_{adj}]=1.14, 95% CI 1.03-1.27, *P*=.01).

As the two subscales (ie, *Behavioral Engagement* and *Experiential Engagement*) were not significantly correlated with each other, an unplanned analysis was conducted to assess the independent association of each subscale with future behavioral engagement. In unadjusted and adjusted analyses, *Experiential Engagement* was significantly associated with future behavioral engagement (OR_{adj}=1.19, 95% CI 1.05-1.34, *P*=.006). In unadjusted and adjusted analyses, *Behavioral Engagement* was not significantly associated with future behavioral engagement (OR_{adj}=1.31, 95% CI 0.38-4.59, *P*=.67).

In unadjusted and adjusted analyses, asking users about how engaging they thought the app was did not significantly predict future behavioral engagement (OR_{adj}=1.34, 95% CI 0.98-1.84, *P*=.07). In unadjusted and adjusted analyses, asking users about how much they liked the app significantly predicted future behavioral engagement (OR_{adj}=1.38, 95% CI 1.03-1.84, *P*=.03).

Table 5. Unadjusted and adjusted odds ratios for the associations between the predictor variables and future behavioral engagement.

Predictor variables	Odds ratio (95% CI)	*P* value	Adjusted odds ratio[a] (95% CI)	*P* value
Total DBCI[b] Engagement Scale score	1.15 (1.05-1.27)	.005	1.14 (1.03-1.27)	.009
Subscale 1 - Experiential Engagement	1.19 (1.06-1.34)	.004	1.19 (1.05-1.34)	.006
Subscale 2 - Behavioral Engagement	1.11 (0.90-1.36)	.34	1.08 (0.87-1.35)	.48
"How engaging was the app?"	1.28 (0.96-1.71)	.097	1.34 (0.98-1.84)	.07
"How much did you like the app?"	1.39 (1.05-1.83)	.02	1.38 (1.03-1.84)	.03

[a]Odds ratios adjusted for motivation to reduce alcohol consumption.

[b]DBCI: digital behavior change intervention.

Incremental Validity

Results from the incremental validity analyses are reported in Table 6. The automatically recorded behavioral indicators of engagement (ie, items 11 and 12; Model 1) accounted for 15.9% of variance in the variable *subsequent login*. The automatically recorded behavioral indicators in combination with the self-reported experiential indicators of engagement (ie, items 1, 2, 3, and 6; Model 2) accounted for 21.1% of variance in the variable *subsequent login*.

Table 6. Odds ratios for the associations between the predictor variables and future behavioral engagement.

Models	Odds ratio (95% CI)	*P* value	Variance accounted for (%)
Model 1			**15.9**
Objectively recorded amount of use	3.46 (1.58-7.57)	.002	
Objectively recorded depth of use	0.91 (0.58-1.42)	.67	
Model 2			**21.1**
Objectively recorded amount of use	2.86 (1.25-6.55)	.013	
Objectively recorded depth of use	0.95 (0.60-1.50)	.82	
Interest	1.72 (1.03-2.85)	.04	
Focus	0.82 (0.50-1.35)	.44	
Enjoyment	0.93 (0.61-1.40)	.72	
Intrigue	1.17 (0.78-1.76)	.45	

Divergent Validity

Total scale scores were significantly correlated with the first ("When using *Drink Less*, the way time passed seemed different from normal") but not the second ("When using *Drink Less*, I was not worried about what others may have been thinking about me") indicator of flow (r_{145}=0.25, *P*<.01 and r_{145}=–0.01, *P*=.95, respectively). The two items tapping flow were not significantly correlated with one another in this sample (r_{145}=–0.06, *P*=.47).

Discussion

Principal Findings

The DBCI Engagement Scale was found to be underpinned by two, largely independent factors, which were labelled

Experiential Engagement and *Behavioral Engagement*. The scale showed moderate internal reliability, but low divergent and criterion validity. Importantly, the behavioral subscale may not be a valid indicator of behavioral engagement. Total scale scores were weakly associated with future behavioral engagement (ie, the variable *subsequent login*), as were the experiential subscale and one of the *best bets* for a short measure of engagement (ie, asking participants about how much they liked the app). The behavioral subscale was not independently associated with future behavioral engagement. In addition, a model including the self-reported experiential and objectively recorded behavioral indicators of engagement (as compared with a model including only the objectively recorded behavioral indicators) accounted for a larger proportion of variance in future behavioral engagement. These findings are at odds with those from the first evaluation of the DBCI Engagement Scale, in which the scale was found to be underpinned by a single factor [14]. However, these differences may at least partly be accounted for by the small sample size in this study.

Construct Validity

The finding that the *Experiential Engagement* and *Behavioral Engagement* subscales were not significantly correlated with each other in this study lends support to the argument that users can spend time on a DBCI without necessarily being interested in or paying attention to its content, and vice versa [14]. However, this finding also gives rise to the question of whether experiential and behavioral engagement are part of the same higher-order construct.

The finding that participants' total scale scores were weakly associated with future behavioral engagement even when adjusting for motivation to reduce alcohol consumption serves as initial evidence that the state of engagement with a DBCI is conceptually distinct from motivation to change the target behavior.

Incremental and Predictive Validity

The results from the incremental validity analyses suggest that behavioral and experiential indicators in tandem have superior predictive power compared with the behavioral indicators alone. However, the finding that the experiential, but not the behavioral, subscale was independently associated with future behavioral engagement can be interpreted to suggest that the experiential indicators (particularly users' interest) were driving the association between initial and future engagement. A potential explanation for these findings is that more intensive engagement during the first login session might have made users' memory of the app more salient, which might have made them more likely to remember to return to the app. As one of the short measures of engagement (ie, the item asking about how much users liked the app) was also found to predict future engagement, it is possible that not only salience of the app, but a salient memory of liking the app, is important for future engagement. It is unclear why the first, but not the second, short measure of engagement had significant predictive power; the word *liking* might be easier to interpret than the word *engaging*. The potential mechanisms underlying the relationship between initial experiential and behavioral engagement, and future behavioral engagement (ie, the variable *subsequent login*) should be explored further using experience sampling techniques in the first few hours following initial app engagement; this involves repeated measurements of psychological processes in real time, in users' natural environments [36].

These results also beg the question as to whether future behavioral engagement is the most appropriate criterion variable to test an engagement scale against. For example, knowledge retention or skill acquisition may be more theoretically sound, as suggested by the Elaboration Likelihood Model of Persuasion (ELMP) [37]. The ELMP argues that deep information processing occurs when an individual pays attention to (or engages with) a health message, which leads to increased knowledge retention. It is plausible that initial behavioral and experiential engagement have superior predictive power compared with behavioral engagement when used to predict knowledge retention. In addition, it would be useful to assess whether the new measure of moment-to-moment (or state-like) engagement is able to predict intervention effectiveness at a later time point.

Criterion Validity

The finding that the self-reported and objectively recorded indicators of *amount of use* were not significantly correlated in this sample suggests that the DBCI Engagement Scale may not be a valid indicator of behavioral engagement. However, although the *amount of use* (ie, time spent in minutes or seconds) is typically used as a gold standard or ground truth of behavioral engagement, our results showed that objectively recorded *amount of use* was significantly correlated with many of the experiential indicators (eg, interest, intrigue). Although the exploratory factor analyses did not indicate that *amount of use* loads onto the same factor as the experiential indicators of engagement, the observed pattern of correlations leads us to question whether time spent on a DBCI is deserving of its ground truth status. There is, hence, a need for future research to investigate the source of the discrepancy between self-reported and objectively recorded indicators of *amount of use*.

Divergent Validity

In line with the first study evaluating the scale, this study did not provide evidence that the DBCI Engagement Scale diverges from the Flow State Scale. There is conceptual overlap between engagement with DBCIs and the dimension of flow that is labelled *losing track of time*. It should be noted that the proposed definition of engagement was, in part, developed based on the concept of flow [14]. It may hence be more fruitful to assess the scale's divergent validity using a more conceptually distinct measure in the future. The lack of evidence that the DBCI Engagement Scale diverges from the Flow State Scale may also serve as a plausible explanation for why participants' self-reported *amount of use* was not significantly correlated with their objectively recorded *amount of use*; they may have lost track of time when engaging with the *Drink Less* app. This finding suggests that self-reported and objectively recorded indicators of time spent on a DBCI may tap different constructs; future research is required to examine which of these is more strongly related to key outcomes of interest.

Limitations

This study was limited because it did not achieve the desired sample size of 250 participants. As Prolific is a novel platform with a small proportion of individuals meeting the study eligibility criteria (ie, drinking alcohol excessively, willing to download an alcohol reduction app, owning an iPhone), the extant participant pool was exhausted after screening just over 400 participants. Although the participant-to-item ratio is considered key in determining the minimum necessary sample size for conducting factor analyses, findings from simulation studies indicate that other factors, including the number of items per factor and the level of communality between items, also influence sample size requirements [38]. Given the limited participant-to-item ratio and the small number of items per factor in this study, the two-factor solution should be interpreted with caution and merits replication in a larger sample in future research. A second limitation is that market research indicates that iOS users are, on average, more affluent than Android users [39]. As the *Drink Less* app is currently available for iOS users only, our findings may not be generalizable to Android users.

Studies conducted via Prolific that involve an initial screening study followed by inviting eligible participants to complete the actual study tend to have attrition rates of approximately 20%-25%, and not 45% [40]. It is therefore likely that there were systematic differences between eligible participants who completed the task and those who did not. For example, the small financial reward may not have been perceived as worth the effort of downloading an app. Indeed, a study assessing the demographic and psychological characteristics of participants who regularly complete research tasks via Amazon's Mechanical Turk online platform (which is similar to Prolific) found that the majority of surveyed participants reported that earning money was a key motivator for taking part [24]. It should also be noted that the financial incentive may have interfered with participants' naturalistic engagement, thus limiting the generalizability of the findings. Previous research has found that money can be an important motivator in DBCI research and increase response rates in longitudinal studies [41].

We did not want to overburden users; hence, we did not assess key trait-like variables that may have influenced users' scale scores. For example, it would have been useful to attempt to partial out the variance accounted for by users' personality traits, such as those specified in the Big Five model of personality [42], to ensure that the DBCI Engagement Scale is detecting something beyond high conscientiousness or low neuroticism.

The adjustment for participants' motivation to reduce their alcohol consumption should have increased the item covariance on the DBCI Engagement Scale and is hence considered a study strength. It should, however, be noted that participants' motivation may have interacted with their engagement levels. Hence, despite the adjustment for participants' motivation to change, the scale scores may not fully represent participants' "true" engagement scores.

Finally, the decision to use Google's cutoff (ie, 30 minutes of inactivity) to identify whether users had made a subsequent login is, to our best knowledge, not grounded in evidence about session length. Future research should explore whether this constitutes a useful heuristic for identifying new DBCI sessions using both quantitative and qualitative methods.

Avenues for Future Research

Due to the observed nonnormal distributions of the scale items that jointly form the DBCI Engagement Scale, a decision was made to use z-score normalization. Consequently, total scores on the DBCI Engagement Scale are only meaningful in relation to the average intensity of experiential and behavioral engagement that a particular DBCI generates. This may facilitate attempts to develop cutoffs for "high" and "low" engagers across DBCIs, irrespective of their specific parameters (eg, the number and length of intervention components). For example, users with scores that fall within a particular range of SDs above or below the mean might usefully be classified as "high" or "low" engagers, and these patterns may replicate across DBCIs. The question of whether the mean and spread of engagement scores replicate across DBCIs merits exploration by evaluating the DBCI Engagement Scale across different kinds of DBCIs (eg, websites or apps for smoking cessation or physical activity).

The finding that initial experiential engagement (or liking of the app) was independently associated with future behavioral engagement suggests that intervention developers should think carefully about how to make their DBCIs appealing on first use. The DBCI Engagement Scale may be useful during the iterative design process, comparing users' experiences of differently designed graphical user interfaces.

Conclusions

The DBCI Engagement Scale assesses behavioral and experiential aspects of engagement. The behavioral subscale may not be a valid indicator of behavioral engagement. The experiential subscale can predict subsequent behavioral engagement with an app for reducing alcohol consumption. Further refinements and validation of the scale in larger samples and across different DBCIs are needed.

Acknowledgments

OP is funded by a PhD studentship from Bupa under its partnership with UCL. CG receives salary support from Cancer Research UK (C1417/A22962). SM is funded by Cancer Research UK and the NIHR School for Public Health Research. We gratefully acknowledge all the funding received.

Authors' Contributions

OP, JL, CG, AB, RW, and SM designed the study. OP collected the data, conducted the statistical analyses, and wrote the first draft of the manuscript. All authors have contributed to the final version of the manuscript and agree with its submission to JMIR.

Conflicts of Interest

OP, CG, AB, and SM have no conflicts of interest to declare. RW undertakes research and consultancy and receives fees for speaking from companies that develop and manufacture smoking cessation medications. JL is an employee at Prolific.

References

1. Donkin L, Christensen H, Naismith S, Neal B, Hickie I, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. J Med Internet Res 2011 Aug 05;13(3):e52 [FREE Full text] [doi: 10.2196/jmir.1772] [Medline: 21821503]
2. Michie S, Yardley L, West R, Patrick K, Greaves F. Developing and Evaluating Digital Interventions to Promote Behavior Change in Health and Health Care: Recommendations Resulting From an International Workshop. J Med Internet Res 2017 Jun 29;19(6):e232 [FREE Full text] [doi: 10.2196/jmir.7126] [Medline: 28663162]
3. Kelders S, Kok R, Ossebaard H, van Gemert-Pijnen JEWC. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012 Nov 14;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]
4. Cobb NK, Graham AL, Bock BC, Papandonatos G, Abrams DB. Initial evaluation of a real-world Internet smoking cessation system. Nicotine Tob Res 2005 Apr;7(2):207-216 [FREE Full text] [doi: 10.1080/14622200500055319] [Medline: 16036277]
5. Alexander GL, McClure JB, Calvi JH, Divine GW, Stopponi MA, Rolnick SJ, et al. A randomized clinical trial evaluating online interventions to improve fruit and vegetable consumption. Am J Public Health 2010 Feb;100(2):319-326 [FREE Full text] [doi: 10.2105/AJPH.2008.154468] [Medline: 20019315]
6. Yardley L, Spring B, Riper H, Morrison L, Crane D, Curtis K, et al. Understanding and Promoting Effective Engagement With Digital Behavior Change Interventions. Am J Prev Med 2016 Nov;51(5):833-842 [FREE Full text] [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
7. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
8. Sieverink F, Kelders SM, van Gemert-Pijnen JE. Clarifying the Concept of Adherence to eHealth Technology: Systematic Review on When Usage Becomes Adherence. J Med Internet Res 2017 Dec 06;19(12):e402 [FREE Full text] [doi: 10.2196/jmir.8578] [Medline: 29212630]
9. Milward J, Drummond C, Fincham-Campbell S, Deluca P. What makes online substance-use interventions engaging? A systematic review and narrative synthesis. Digit Health 2018;4:2055207617743354 [FREE Full text] [doi: 10.1177/2055207617743354] [Medline: 29942622]
10. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
11. Bellg AJ, Borrelli B, Resnick B, Hecht J, Minicucci DS, Ory M, et al. Enhancing Treatment Fidelity in Health Behavior Change Studies: Best Practices and Recommendations From the NIH Behavior Change Consortium. Health Psychology 2004;23(5):443-451. [doi: 10.1037/0278-6133.23.5.443]
12. Csikszentmihalyi M. Flow: The psychology of optimal performance. New York: Cambridge University Press; 1990.
13. Brown E, Cairns P. ACM Digital Library. 2004. A grounded investigation of game immersion. URL: https://dl.acm.org/citation.cfm?doid=985921.986048 [accessed 2019-11-12]
14. Perski O, Blandford A, Garnett C, Crane D, West R, Michie S. A self-report measure of engagement with digital behavior change interventions (DBCIs): development and psychometric evaluation of the "DBCI Engagement Scale". Transl Behav Med 2019 Mar 30. [doi: 10.1093/tbm/ibz039] [Medline: 30927357]
15. Short CE, DeSmet A, Woods C, Williams SL, Maher C, Middelweerd A, et al. Measuring Engagement in eHealth and mHealth Behavior Change Interventions: Viewpoint of Methodologies. J Med Internet Res 2018 Nov 16;20(11):e292. [doi: 10.2196/jmir.9397]
16. Pham Q, Graham G, Carrion C, Morita PP, Seto E, Stinson JN, et al. A Library of Analytic Indicators to Evaluate Effective Engagement with Consumer mHealth Apps for Chronic Conditions: Scoping Review. JMIR Mhealth Uhealth 2019 Jan 18;7(1):e11941. [doi: 10.2196/11941]
17. Miller S, Ainsworth B, Yardley L, Milton A, Weal M, Smith P, et al. A Framework for Analyzing and Measuring Usage and Engagement Data (AMUsED) in Digital Interventions: Viewpoint. J Med Internet Res 2019 Feb 15;21(2):e10966. [doi: 10.2196/10966]

18. Murray E, White IR, Varagunam M, Godfrey C, Khadjesari Z, McCambridge J. Attrition revisited: adherence and retention in a web-based alcohol trial. J Med Internet Res 2013;15(8):e162 [FREE Full text] [doi: 10.2196/jmir.2336] [Medline: 23996958]
19. Postel M, de Haan HA, ter Huurne ED, van der Palen J, Becker E, de Jong CAJ. Attrition in web-based treatment for problem drinkers. J Med Internet Res 2011 Dec 27;13(4):e117 [FREE Full text] [doi: 10.2196/jmir.1811] [Medline: 22201703]
20. Radtke T, Ostergaard M, Cooke R, Scholz U. Web-Based Alcohol Intervention: Study of Systematic Attrition of Heavy Drinkers. J Med Internet Res 2017 Jun 28;19(6):e217 [FREE Full text] [doi: 10.2196/jmir.6780] [Medline: 28659251]
21. Perski O, Lumsden J, Garnett C, Blandford A, West R, Michie S. Second Evaluation of the DBCI Engagement Scale. Open Sci Framew 2019 [FREE Full text]
22. Babor T, Higgins-Biddle J, Saunders J, Monteiro M. The Alcohol Use Disorders Identification Test: Guidelines for Use in Primary Care. 2nd ed. Geneva: World Health Organization; 2001.
23. Peer E, Brandimarte L, Samat S, Acquisti A. Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. Journal of Experimental Social Psychology 2017 May;70:153-163. [doi: 10.1016/j.jesp.2017.01.006]
24. Paolacci G, Chandler J, Ipeirotis P. Running experiments on Amazon Mechanical Turk. Judgment and Decision Making 2010;5(5):411-419 [FREE Full text] [Medline: 1011330798]
25. Costello A, Osborne JW. Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. Practical Assessment, Research & Evaluation 2005;10:173-178.
26. Kotz D, Brown J, West R. Predictive validity of the Motivation To Stop Scale (MTSS): a single-item measure of motivation to stop smoking. Drug Alcohol Depend 2013 Feb 01;128(1-2):15-19 [FREE Full text] [doi: 10.1016/j.drugalcdep.2012.07.012] [Medline: 22943961]
27. Hummel K, Brown J, Willemsen MC, West R, Kotz D. External validation of the Motivation To Stop Scale (MTSS): findings from the International Tobacco Control (ITC) Netherlands Survey. Eur J Public Health 2017 Feb 01;27(1):129-134. [doi: 10.1093/eurpub/ckw105] [Medline: 28177479]
28. de Vocht F, Brown J, Beard E, Angus C, Brennan A, Michie S, et al. Temporal patterns of alcohol consumption and attempts to reduce alcohol intake in England. BMC Public Health 2016 Sep 01;16:917 [FREE Full text] [doi: 10.1186/s12889-016-3542-7] [Medline: 27585991]
29. Crane D, Garnett C, Michie S, West R, Brown J. A smartphone app to reduce excessive alcohol consumption: Identifying the effectiveness of intervention components in a factorial randomised control trial. Sci Rep 2018 Mar 12;8(1):4384 [FREE Full text] [doi: 10.1038/s41598-018-22420-8] [Medline: 29531280]
30. Google Analytics Help. 2017. How a web session is defined in Analytics. URL: https://support.google.com/analytics/answer/2731565 [accessed 2018-02-06]
31. Braze Magazine. Spring 2016 Mobile Customer Retention Report. 2016. URL: https://www.braze.com/blog/app-customer-retention-spring-2016-report/ [accessed 2019-09-09]
32. CISION PRWeb. 2015. Motivating Patients to Use Smartphone Health Apps. URL: http://www.prweb.com/releases/2011/04/prweb5268884.htm [accessed 2015-08-10]
33. Jackson S, Marsh H. Development and validation of a scale to measure optimal experience: The Flow State Scale. J Sport Exerc Psychol 1996;18:35 [FREE Full text]
34. Wiebe E, Lamb A, Hardy M, Sharek D. Measuring engagement in video game-based environments: Investigation of the User Engagement Scale. Comput Human Behav 2014;32:132. [doi: 10.1016/j.chb.2013.12.001]
35. Hinkin T. A Brief Tutorial on the Development of Measures for Use in Survey Questionnaires. Organ Res Methods 1998;1(1):121. [doi: 10.1177/109442819800100106]
36. Stone A, Shiffman S. Ecological Momentary Assessment (Ema) in Behavioral Medicine. Annals of Behavioral Medicine 1994;16(3):199-202. [doi: 10.1093/abm/16.3.199]
37. Petty R, Cacioppo J. The Elaboration Likelihood Model of Persuasion. Adv Exp Soc Psychol 1986;19:205. [doi: 10.1016/S0065-2601(08)60214-2]
38. Mundfrom DJ, Shaw DG, Ke TL. Minimum Sample Size Recommendations for Conducting Factor Analyses. International Journal of Testing 2005;5(2):159-168. [doi: 10.1207/s15327574ijt0502_4]
39. Berg M. Statista. 2014. iPhone Users Earn More. URL: https://www.statista.com/chart/2638/top-line-platform-stats-for-app-usage-in-the-us/ [accessed 2019-01-28]
40. Palan S, Schitter C. Prolific.ac—A subject pool for online experiments. Journal of Behavioral and Experimental Finance 2018 Mar;17:22-27 [FREE Full text] [doi: 10.1016/j.jbef.2017.12.004]
41. Khadjesari Z, Murray E, Kalaitzaki E, White I, McCambridge J, Thompson S, et al. Impact and costs of incentives to reduce attrition in online trials: two randomized controlled trials. J Med Internet Res 2011 Mar 02;13(1):e26 [FREE Full text] [doi: 10.2196/jmir.1523] [Medline: 21371988]
42. John OP, Donahue EM, Kentle RL. The Big Five Inventory. 1991. URL: http://www.sjdm.org/dmidi/Big_Five_Inventory.html [accessed 2019-09-09]

Abbreviations

AUDIT: Alcohol Use Disorders Identification Test
DBCI: digital behavior change intervention

Edited by G Eysenbach; submitted 09.09.19; peer-reviewed by S Miller, J Thrul; comments to author 17.10.19; revised version received 30.10.19; accepted 11.11.19; published 20.11.19.

Please cite as:
Perski O, Lumsden J, Garnett C, Blandford A, West R, Michie S
Assessing the Psychometric Properties of the Digital Behavior Change Intervention Engagement Scale in Users of an App for Reducing Alcohol Consumption: Evaluation Study
J Med Internet Res 2019;21(11):e16197
URL: http://www.jmir.org/2019/11/e16197/
doi:10.2196/16197
PMID:31746771

Viewpoint

Implementation: The Next Giant Hurdle to Clinical Transformation With Digital Health

Lorraine Buis[1], PhD, MSI

Department of Family Medicine, University of Michigan, Ann Arbor, MI, United States

Corresponding Author:
Lorraine Buis, PhD, MSI
Department of Family Medicine
University of Michigan
1018 Fuller St
Ann Arbor, MI, 48104
United States
Phone: 1 734 998 7120
Email: buisl@umich.edu

Abstract

Clinical implementation of digital health is a major hurdle to overcome in the coming years. Considering the role of the *Journal of Medical Internet Research* in the past 20 years and looking toward the journal's future, this viewpoint acknowledges the vision of medicine and the role that digital health plays in that vision. It also highlights barriers to implementation of digital health as an obstacle to achieving that vision. In particular, this paper focuses on how digital health research must start looking toward implementation as an area of inquiry and the role that the *Journal of Medical Internet Research* and its' sister journals from JMIR Publications can play in this process.

(J Med Internet Res 2019;21(11):e16259) doi:10.2196/16259

KEYWORDS

digital health; digital medicine; ehealth; mhealth; implementation; knowledge translation; publishing; open access; journalogy

Digital health is quickly revolutionizing the way health care is conducted. For two decades, the *Journal of Medical Internet Research* (JMIR) has been the premier outlet for publishing digital health research, and as the field has grown more sophisticated, so has the research itself. Starting with studies of health-related websites and evolving to online communities with message boards, Web-based interventions, social networking sites, mobile health interventions, and remote sensing, JMIR has helped build the evidence base, from descriptive studies to randomized controlled trials.

As we take a moment to reflect on where we have been and look to the future to see where we are going, the shared vision of health care is clear. Without a doubt, digital health is here to stay. Digital technologies supporting the collection of patient-generated health data, and algorithms and tools for interpreting and managing the copious amounts of data, will serve to put real-world data into the hands of health care providers who can act on it in real time. Digital health has the potential to connect individuals to patients, providers, and data; assist with self-management and behavior change; level the playing field to reduce health disparities and improve access to care; and surveil and disseminate information for public health, including in times of disaster. If all goes according to plan, digital health will accomplish these aims as well as improve efficiency in both clinical workflow and cost. Although this future vision of health care is anticipated and hoped for by many, and ongoing work today is demonstrating the feasibility of digital health in accomplishing these aims, the road to how we achieve this vision still unknown.

To illustrate, one question that has emerged in my own work is how to fit effective digital health interventions into existing clinical practice. As researchers in an emerging field, we tend to put technology into the hands of doctors and nurses to ask a research question about how to improve health outcomes. There are thousands of studies in the literature reporting feasibility, acceptability, and effectiveness of digital health interventions. However, we do not have the answers to how we could make effective technologies practicable in clinical settings, which is the piece needed to make our future vision of health care a reality.

With roots in the area of diffusion of innovations, popularized by Everett Rogers in the 1960s [1], the focus on dissemination and implementation research within health care contexts has been rapidly gaining momentum over the past few decades [2]. Although these concepts are not new, the increasing focus on

factors that affect translation of evidence-based interventions, practices, and policies into real-world settings within health care has brought dissemination and implementation research to the forefront of our national research conversation. It is increasingly understood that it is not enough to just push a new development into the world and hope it sticks. Regarding digital health, researchers are just starting to look at implementation strategies themselves to find efficient ways to enhance implementation efforts [3-5] as well as to evaluate digital health implementations to understand why interventions perform the way they do [6]. Naturally, questions arise, such as how do we fit these technologies into existing workflow practices where providers do not have the time or training to utilize them effectively within clinical practice? Who will serve to troubleshoot problems experienced by providers and patients? Who will help patients, particularly less savvy ones, get up and running? How will providers fit the use of these technologies into already packed clinical encounters? How will health systems pay for these technologies? What are the ethical considerations that must be made before widespread dissemination can occur? Issues such as these are where the rubber meets the road and will make or break our realization of a shared vision of health care. Without answers to these questions, our vision is just a dream.

To meet the demands of our future needs related to digital health, one implementation strategy might be to have existing medical professionals such as medical assistants, nurses, physician assistants, or others shift roles to include new tasks created by the introduction of new technologies within health care. However, it is possible that if we were to fast forward 20 years, we may see an entirely new class of health care providers: tech-savvy clinical professionals who assist with patient-generated health data and remote sensing, are specially trained to monitor dashboards of patient-generated health data, and are assisted by algorithms to detect issues related to safety. These professionals, embedded in the health care systems they serve, would be uniquely poised to communicate with care teams about changes in conditions and recommend acute or routine follow-up. Much like the introduction of medical scribes within clinical care as a way to manage documentation burdens imposed by electronic health record systems [7], specially trained personnel who can help manage burdens associated with the collection of patient-generated health data may be useful.

Although it is intuitive that the development of a role such as this would be helpful, in the current publishing landscape, it is unclear what publication outlets would consider research into this type of implementation strategy to be in scope. As is often the case with this kind of multidisciplinary work, this work could have many possible homes including journals on practice management, clinical informatics, medicine, nursing, and public health; however, it does not squarely fit the mold of what one would expect in discipline-specific publications. This is exactly why JMIR (and its sister journals) has become, and will continue to be, the pre-eminent home for research on digital health. JMIR has been publishing manuscripts related to focused implementation of digital health technologies for two decades. From papers focused on lessons learned [8] to mixed methods studies looking for a deeper understanding of digital health in practice [9], studies of which delivery modality works best [10], and understanding characteristics of digital health among nonusers/nonadopters and users/adopters and reasons for continued engagement or dropout to gain a deeper understanding [11-17], JMIR journals have been a welcoming home for this work.

In the coming years, we will continue to have the need for research detailing efficacy and effectiveness; however, more research exploring optimal configuration of interventions, implementation strategies, cutting-edge evaluation methods for digital health, and more widely focused research centered on complex implementation issues is certainly on the horizon. As we look toward the next two decades and beyond, implementation issues are paramount, and through dissemination in one of the many journals within the JMIR family or a new JMIR publication yet to be developed, implementation will surely become a critical topic of interest to the readers, yielding yet another opportunity for JMIR to continue leading the way.

Conflicts of Interest

None declared.

References

1. Rogers EM. Diffusion of innovations. New York: The Free Press of Glencoe; 1962.
2. Norton WE, Lungeanu A, Chambers DA, Contractor N. Mapping the growing discipline of dissemination and implementation science in health. Scientometrics 2017 Jul 11;112(3):1367-1390. [doi: 10.1007/s11192-017-2455-2]
3. Varsi C, Solberg Nes L, Kristjansdottir OB, Kelders SM, Stenberg U, Zangi HA, et al. Implementation Strategies to Enhance the Implementation of eHealth Programs for Patients With Chronic Illnesses: Realist Systematic Review. J Med Internet Res 2019 Sep 27;21(9):e14255 [FREE Full text] [doi: 10.2196/14255] [Medline: 31573934]
4. Quanbeck A. Using Stakeholder Values to Promote Implementation of an Evidence-Based Mobile Health Intervention for Addiction Treatment in Primary Care Settings. JMIR Mhealth Uhealth 2019 Jun 07;7(6):e13301. [doi: 10.2196/13301]
5. East ML, Havard BC. Mental Health Mobile Apps: From Infusion to Diffusion in the Mental Health Social System. JMIR Mental Health 2015 Mar 31;2(1):e10. [doi: 10.2196/mental.3954]
6. Ware P, Ross HJ, Cafazzo JA, Laporte A, Gordon K, Seto E. Evaluating the Implementation of a Mobile Phone-Based Telemonitoring Program: Longitudinal Study Guided by the Consolidated Framework for Implementation Research. JMIR Mhealth Uhealth 2018 Jul 31;6(7):e10768 [FREE Full text] [doi: 10.2196/10768] [Medline: 30064970]

7. Campbell LL, Case D, Crocker JE, Foster M, Johnson M, Lee CA, et al. Using medical scribes in a physician practice. J AHIMA 2012;83(11):64-69. [Medline: 23210302]
8. Mechael P, Kaonga NN, Chandrasekharan S, Prakash MP, Peter J, Ganju A, et al. The Elusive Path Toward Measuring Health Outcomes: Lessons Learned From a Pseudo-Randomized Controlled Trial of a Large-Scale Mobile Health Initiative. JMIR Mhealth Uhealth 2019 Aug 21;7(8):e14668. [doi: 10.2196/14668]
9. Tunnecliff J, Ilic D, Morgan P, Keating J, Gaida JE, Clearihan L, et al. The Acceptability Among Health Researchers and Clinicians of Social Media to Translate Research Evidence to Clinical Practice: Mixed-Methods Survey and Interview Study. J Med Internet Res 2015 May 20;17(5):e119. [doi: 10.2196/jmir.4347]
10. Narayanaswami P, Gronseth G, Dubinsky R, Penfold-Murray R, Cox J, Bever Jr C, et al. The Impact of Social Media on Dissemination and Implementation of Clinical Practice Guidelines: A Longitudinal Observational Study. J Med Internet Res 2015 Aug 13;17(8):e193. [doi: 10.2196/jmir.4414]
11. Bouwman T, van Tilburg T, Aartsen M. Attrition in an Online Loneliness Intervention for Adults Aged 50 Years and Older: Survival Analysis. JMIR Aging 2019 Jul 24;2(2):e13638. [doi: 10.2196/13638]
12. Duggleby W, Ploeg J, McAiney C, Fisher K, Jovel Ruiz K, Ghosh S, et al. A Comparison of Users and Nonusers of a Web-Based Intervention for Carers of Older Persons With Alzheimer Disease and Related Dementias: Mixed Methods Secondary Analysis. J Med Internet Res 2019 Oct 17;21(10):e14254 [FREE Full text] [doi: 10.2196/14254] [Medline: 31625947]
13. Grutzmacher SK, Munger AL, Speirs KE, Vafai Y, Hilberg E, Braunscheidel Duru E, et al. Predicting Attrition in a Text-Based Nutrition Education Program: Survival Analysis of Text2BHealthy. JMIR Mhealth Uhealth 2019 Jan 21;7(1):e9967. [doi: 10.2196/mhealth.9967]
14. Kamke K, Grenen E, Robinson C, El-Toukhy S. Dropout and Abstinence Outcomes in a National Text Messaging Smoking Cessation Intervention for Pregnant Women, SmokefreeMOM: Observational Study. JMIR Mhealth Uhealth 2019 Oct 07;7(10):e14699 [FREE Full text] [doi: 10.2196/14699] [Medline: 31593542]
15. Sanders I, Short CE, Bogomolova S, Stanford T, Plotnikoff R, Vandelanotte C, et al. Characteristics of Adopters of an Online Social Networking Physical Activity Mobile Phone App: Cluster Analysis. JMIR Mhealth Uhealth 2019 Jun 03;7(6):e12484. [doi: 10.2196/12484]
16. Vaghefi I, Tulu B. The Continued Use of Mobile Health Apps: Insights From a Longitudinal Study. JMIR Mhealth Uhealth 2019 Aug 29;7(8):e12983. [doi: 10.2196/12983]
17. Wiseman KP, Coa KI, Prutzman YM. Predictors of Retention in an Adult Text Messaging Smoking Cessation Intervention Program: Cohort Study. JMIR Mhealth Uhealth 2019 Aug 01;7(8):e13712 [FREE Full text] [doi: 10.2196/13712] [Medline: 31373278]

Abbreviations

JMIR: Journal of Medical Internet Research

Edited by G Eysenbach; submitted 13.09.19; peer-reviewed by C Harshbarger; comments to author 13.10.19; revised version received 22.10.19; accepted 11.11.19; published 20.11.19.

Please cite as:
Buis L
Implementation: The Next Giant Hurdle to Clinical Transformation With Digital Health
J Med Internet Res 2019;21(11):e16259
URL: http://www.jmir.org/2019/11/e16259/
doi:10.2196/16259
PMID:31746763

Viewpoint

Digital Health Consumers on the Road to the Future

Rita Kukafka[1,2], DrPH, MA

[1]Department of Biomedical Informatics, Columbia University, New York, NY, United States
[2]Department of Sociomedical Sciences, Columbia University, New York, NY, United States

Corresponding Author:
Rita Kukafka, DrPH, MA
Department of Biomedical Informatics
Columbia University
622 West 168th Street
PH-20
New York, NY, 10032
United States
Phone: 1 2123059193
Email: rk326@cumc.columbia.edu

Abstract

Digital health is uniquely positioned to transform health care. This viewpoint explores the enormous benefits for health consumers when digital-first health care is embraced. Also, it explores what risks exist if surveillance capitalism takes over health care. Further, some solutions to prepare digital health citizens for the road ahead are also discussed.

(J Med Internet Res 2019;21(11):e16359) doi:10.2196/16359

KEYWORDS

digital health; health consumers; artificial intelligence; internet of things

Introduction

Digital health is uniquely positioned to transform health care. Ubiquitous computing and the technological advancements of mobile-computing platforms and wearable consumer devices have enabled continuous monitoring of citizens and their everyday behaviors. This longitudinal data can be mined to disclose physiological and behavioral signatures of existing health impairments and may effectively provide predictive capabilities for health conditions yet to emerge. The most complete multi-modal data provide the best predictive capabilities, which then yield powerful pathways for preventing disease. Digital health consumers who provide the data have the most to gain, but they also have the potential for loss. This viewpoint explores the evolving lexicon and landscape of digital health, as well as some risks and benefits digital health consumers may face in the future. It also explores what risks exist if surveillance capitalism takes over health care. Finally, some solutions to prepare digital health citizens for the road ahead are discussed.

An Evolving Lexicon of Terms for Digital Transformation

Digital health is the broad umbrella term which, according to the World Health Organization (WHO), encompasses electronic health (eHealth) but has also been broadened to include connected software solutions (Internet of Things [IoT]), as well as computational methods applied to big data, genomics, and artificial intelligence (AI) [1]. Agreement on how to define related terms has been brewing for more than a decade (eHealth, medical informatics, telemedicine, and mobile health) [2-4], and a key benefit to proclaiming an umbrella term stems from the need to integrate related concepts that are agnostic to specificities, such as purpose and type of technology. Noteworthy, the *Journal of Medical Internet Research* (JMIR) advanced the need for a broad, unifying term decades ago [2], and early on embraced the ubiquitous and transformative nature of digital technologies:

> *e-health is an emerging field in the intersection of medical informatics, public health and business, referring to health services and information delivered or enhanced through the Internet and related technologies. In a broader sense, the term characterizes not only a technical development, but*

> *also a state-of-mind, a way of thinking, an attitude, and a commitment for networked, global thinking, to improve health care locally, regionally, and worldwide by using information and communication technology.*

Regarding the expanding lexicon of terms, two points pertinent to the future path of digital consumers can be made. First, the quest to redefine terms will naturally occur as the field emerges, typically as a result of interdisciplinary, multidisciplinary, and transdisciplinary interactions of individuals and concepts in a way that transcends conventional field boundaries [5]. Nonetheless, it is the preservation of information contained in thousands of published articles and lessons learned that remain the core for understanding digital health consumers moving forward. The second, interrelated point also pertains to field boundaries and provides a clue to a future road. The WHO's declaration to designate digital health as an umbrella term to encompass and broaden eHealth does not arise from expansion to new domains or of novel digital health technologies. Instead, it reflects the convergence of the eHealth landscape with fields such as AI, big data, and genomics. Big data, IoT, advances in computing power, and memory collectively set the stage for the future, enabling a remarkable unchartered path for digital health consumers and our conception of the space they encounter. It is at this intersection that the eHealth landscape is therefore positioned to advance and transform health care.

Consumers' Evolving Use of Digital Technologies

Just as the lexicon of terms has evolved to reflect expanding boundaries, so has the digital health consumer's interactions with technologies also evolved. Consumer adoption of digital technologies has extensively increased [6], and it is also of note that consumers are evolving in their use of digital technologies [7]. However, as reflected in the preceding discussion, the most resounding change for digital consumers today is that they now reside amid a network of smart things that are capable of capturing their digital footprints. These smart things come with embedded sensors and wireless devices and form a state of extreme digital connectivity. The IoT first described by Kevin Ashton in 1999 describes this as a network of physical things that can sense other things using ubiquitous wireless connectivity and embedded sensors [8]. Ubiquitous and wireless connectivity is expected to grow in future years.

Today we live on a planet that has more smart things than human beings. The global IoT health care market, valued at US $5800 million in 2014, is expected to reach a value of approximately US $14,000 million by 2024 [9]. Around 29 billion connected devices are forecast by 2022, of which around 18 billion will be related to IoT [10]. The IoT streams in big data from smart things and can virtually and digitally connect inanimate and physical living things. With speed and efficiency, AI strategies are ideally adapted to managing and analyzing these continuous data streams in large amounts [11]. Unparalleled conveniences, benefits, and challenges persist at the intersection of IoT, AI, and health care [12]. If consumers are currently overwhelmed about how best to manage the growing number of smart objects at work and in everyday life, it is interesting to note that by 2030, each person will own 15 connected devices [13]. We are only experiencing the beginning of a fast-approaching deluge of smart things.

The Road Ahead for Digital Health Consumers

In academic literature, digital health interventions are commonly presented as ways to deepen patient activation or involvement in care, promote behavior change, and harness or promote self-management [14]. It was during the 1990s, as the internet exploded into public consciousness, that several eHealth terms began to emerge [2], and consumer use of the internet for health information reached exponential growth [15]. Both the emergence of Web 2.0 technologies and personal health applications, such as Google Health and Microsoft HealthVault, were viewed as having far-reaching consequences for patient involvement [16,17]. It was thought that health care would become democratized if patients had access to their medical records and access to the medical literature. Lay people would become more conversant with health and medical issues, and the traditional paternalistic paradigm was expected to shift. The IoT for digital health consumers may be the next evolution of the current internet into a network of interconnected smart things that not only gather information from the environment (sensing) and interact with the physical world, but also use existing internet standards to provide services for information transfer, analytics, application, and communication [18]. So how is the road forward expected to be different for digital consumers living in this world of ubiquitous and wireless connectivity?

In a highly connected, digitized world, more emphasis will be placed on the digital consumer. Rather than viewing the electronic health record (EHR) as the sanctified knowledge resource for each patient, data will be generated in diverse spaces and places (IoT), including the real world where patients work and live. There will likely be more information-sharing relationships: one-to-many, many-to-one, and many-to-any. This shift is also expected to improve the stickiness of smart devices, as providers get more value from data to monitor patients and as consumers become further incentivized to sustain use [7]. Smartwatches, smartphones, and an array of ubiquitous devices with embedded identification, sensing, and data exchange features enable data gathering in real-time for synchronous communications, personalized interventions, and monitoring patient processes and outcomes. Collectively, the IoT offers considerable potential for transforming health care from encounter-based care into connected continuous care [9]. These rapid changes that are interwoven into health care expand its capabilities and have only just begun to be documented and analyzed in the academic, critical, social-scientific literature.

Topol, in his book "Deep Medicine", describes the virtual medical coach that is designed to support the needs of health consumers as the unfulfilled promise of big data, deep learning, and other AI tools [19]. One of the most important boons of deep medicine:

> *...is to empower not just physicians to be better at what they do, but to help all of us be as good as we can be at taking care of our own health.*

The potential of the virtual medical coach would only be realized if it was merged with behavioral science methods to promote behavior change since so much of the burden of disease is related to poor lifestyle. As an example, IoT surveillance combined with individual sensors from smart devices and other relevant data might trigger behavioral nudges designed to address an impending asthma attack, rising blood pressure, increased BMI, or elevated glucose levels.

The inaugural editorial welcoming readers to JMIR, written by Gunther Eysenbach, highlighted the important role the internet plays in consumer empowerment and in redefining the traditional model of preventive medicine and health promotion. Since then, hundreds of JMIR's papers have focused on participatory approaches to increase the health care experience for consumers and to empower them to take care of their health. The Journal of Participatory Medicine (JoPM), now published by JMIR Publications, pertains specifically to participatory medicine, a movement in which patients and health professionals actively collaborate and encourage one another as full partners in care. JMIR Publications, which is designed to explore the latest research in the field of digital health, provides a critical resource for the continued expansion of digital health technologies as the eHealth landscape continues to converge with the IoT, AI, big data, and genomics.

Risks and Benefits of Electronic Health for Digital Health Consumers

There have been some promising examples of eHealth applications coupled with AI and the IoT. In the area of health monitoring and risk prediction, AI can use raw data from sensors and machine-learning algorithms can then be trained to recognize patterns from the raw data inputs. Patterns can then be categorized as indicators of an individual's behavior and health status, which allow patients to understand and manage their health as well as share data with medical providers. For example, a recent literature review published in JMIR analyzed the literature from 2010 to 2018 and yielded 1849 pertinent articles that combined AI with the latest technologies for diabetes management and decision support [20]. Ultimately, 141 papers were included in the review, which demonstrated the potential of AI to enable diabetes solutions in the context of multiple critical management issues, including blood glucose prediction and control, detection of adverse glycemic events and risk, and patient personalization. In the area of precision medicine, eHealth combined with IoT and AI can be used not only to predict outcomes but also to predict future patients' probability of having diseases, including the probability of not having a disease [21].

As IoT and AI in health care converge to be transformative for the digitally engaged consumer, there are also new risks and challenges. Data for its own sake does not have much value. It is only when information and knowledge are derived from data that actionable utility becomes possible, however, the end-users of this data may vary enormously. This begs the question: what happens when the "underlying model of improved health care and consumer empowerment" becomes "control of humans for profit?" Zuboff raises concerns in her book "The Age of Surveillance Capitalism" about what she calls behavioral futures and prediction products [22]. According to Zuboff, surveillance capitalism: unilaterally claims human experience as free raw material for translation into behavioral data. Although some of these data are applied to service improvement, the rest are declared as a proprietary behavioral surplus, fed into advanced manufacturing processes known as "machine intelligence," and fabricated into prediction products that anticipate what you will do now, soon and later.

Digital health consumers may stand more to lose than their privacy. What is at stake is their autonomy and right to make their own decisions about their health. Nudging health consumers based on AI behavioral predictions may seem praiseworthy when the end-user goal is to improve health, but what happens when consumers are nudged solely to purchase specific products, seek a specific treatment, or vote on a specific piece of health care legislation? From a health care viewpoint, this begs the question of how consumer advocacy groups could take control of digital health from corporate and political interests so that benefits could be gained without the harms that could result from an algorithmic technocracy concerned with profit and control.

Indeed, there are many other challenges digital health consumers will face on their road to the future. Digital consumers may be comfortable relying on AI when using Uber and Amazon, but that is different than relying on AI to guide their medical treatment plans. This concern is justifiable considering that, of the millions of digital health devices available for consumers today, only a small fraction have been tested and the evidence for those that had been evaluated was of low quality [23,24]. Smart and not so smart devices may generate diagnoses, make recommendations, and in some instances, may be authorized to act [25]. Current Health Insurance Portability and Accountability Act (HIPAA) regulations, specifically the Security Rule, discuss the accessibility, integrity, and confidentially of all electronic, protected, health information, but they do not specifically govern IoT devices. Many unanswered questions highlight the need for additional legislation dictating who is responsible for the protection of IoT, such as, does an app or IoT device manufacturer owe consumers a HIPPA level of security for maintaining records of health insights? It is expected that the IoT industry may face additional regulations by the Food and Drug Administration, the US Department of Health and Human Services Office of Civil Rights, and other entities, plus health care–specific regulations.

Calls to action have been effective at spreading the message that digital health will bring patient empowerment, while others have been encouraging a more balanced view that seeks to capitalize on the benefits of digital health while minimizing the risks of potential harms [26-28].

Conclusions

We have only just begun to scrape at the ethical implications for future digital health consumers. It is often challenging to

assess who the main drivers of this “creative destruction” are [29]. There are so many stakeholders, most of them from nonhealth conglomerates, and many of them transnational. The speed of development is surprisingly rapid, and since we are part of the change it is hard to step back to see precisely where we are going. Predicting human behavior to empower digital health consumers to prevent disease may not be as profitable as trading behavioral predictions for profit. Those of us working in academia at the intersection of digital health, ethics, and social justice must be aware and stay ahead of the challenges digital health consumers will face in the future. The road digital health consumers take into the future may be very different from roads taken in the past.

Conflicts of Interest

None declared.

References

1. World Health Organization. 2019. WHO guideline: recommendations on digital health interventions for health system strengthening URL: https://www.who.int/reproductivehealth/publications/digital-interventions-health-system-strengthening/en/ [accessed 2019-09-14]
2. Eysenbach G. What is e-health? J Med Internet Res 2001 Jun 18;3(2):e20 [FREE Full text] [doi: 10.2196/jmir.3.2.e20] [Medline: 11720962]
3. Boogerd EA, Arts T, Engelen LJ, van de Belt TH. "What Is eHealth": Time for An Update? JMIR Res Protoc 2015 Mar 12;4(1):e29 [FREE Full text] [doi: 10.2196/resprot.4065] [Medline: 25768939]
4. Oh H, Rizo C, Enkin M, Jadad A. What is eHealth (3): a systematic review of published definitions. J Med Internet Res 2005 Feb 24;7(1):e1 [FREE Full text] [doi: 10.2196/jmir.7.1.e1] [Medline: 15829471]
5. Casadevall A, Fang FC. Field Science—the Nature and Utility of Scientific Fields. mBio 2015 Sep 08;6(5):1-4. [doi: 10.1128/mbio.01259-15]
6. Robbins R, Krebs P, Jagannathan R, Jean-Louis G, Duncan DT. Health App Use Among US Mobile Phone Users: Analysis of Trends by Chronic Disease Status. JMIR Mhealth Uhealth 2017 Dec 19;5(12):e197 [FREE Full text] [doi: 10.2196/mhealth.7832] [Medline: 29258981]
7. Day S, Zweig M. Rock Health. Beyond Wellness for the Healthy: Digital Health Consumer Adoption 2018 URL: https://rockhealth.com/reports/beyond-wellness-for-the-healthy-digital-health-consumer-adoption-2018/ [accessed 2019-11-14]
8. Gabbai A. Smithsonian.com. 2015 Jan. Kevin Ashton Describes “the Internet of Things” URL: https://www.smithsonianmag.com/innovation/kevin-ashton-describes-the-internet-of-things-180953749/ [accessed 2019-09-05]
9. Meinert E, Van Velthoven M, Brindley D, Alturkistani A, Foley K, Rees S, et al. The Internet of Things in Health Care in Oxford: Protocol for Proof-of-Concept Projects. JMIR Res Protoc 2018 Dec 04;7(12):e12077 [FREE Full text] [doi: 10.2196/12077] [Medline: 30514695]
10. The Ericson Report. Internet of Things Outlook URL: https://www.ericsson.com/en/mobility-report/internet-of-things-outlook
11. Özdemir V. The Big Picture on the "AI Turn" for Digital Health: The Internet of Things and Cyber-Physical Systems. OMICS 2019 Jun;23(6):308-311. [doi: 10.1089/omi.2019.0069] [Medline: 31066623]
12. Fogel AL, Kvedar JC. Artificial intelligence powers digital medicine. Digit Med 2018 Mar 14;1(1):5.
13. Heslop B. Martech Advisor. 2019 Mar 04. By 2030, each person will own 15 connected devices-here's what that means for your business and content URL: https://www.martechadvisor.com/articles/iot/by-2030-each-person-will-own-15-connected-devices-heres-what-that-means-for-your-business-and-content/ [accessed 2019-09-11]
14. Michie S, Yardley L, West R, Patrick K, Greaves F. Developing and Evaluating Digital Interventions to Promote Behavior Change in Health and Health Care: Recommendations Resulting From an International Workshop. J Med Internet Res 2017 Jun 29;19(6):e232 [FREE Full text] [doi: 10.2196/jmir.7126] [Medline: 28663162]
15. Cline R, Haynes K. Consumer health information seeking on the Internet: the state of the art. Health Educ Res 2001 Dec;16(6):671-692. [doi: 10.1093/her/16.6.671] [Medline: 11780707]
16. Eysenbach G. Medicine 2.0: social networking, collaboration, participation, apomediation, and openness. J Med Internet Res 2008 Aug 25;10(3):e22 [FREE Full text] [doi: 10.2196/jmir.1030] [Medline: 18725354]
17. Mandl KD, Kohane IS. Tectonic Shifts in the Health Information Economy. N Engl J Med 2008 Apr 17;358(16):1732-1737. [doi: 10.1056/nejmsb0800220]
18. Gubbi J, Buyya R, Marusic S, Palaniswami M. Internet of Things (IoT): A vision, architectural elements, and future directions. Future Generation Computer Systems 2013 Sep;29(7):1645-1660. [doi: 10.1016/j.future.2013.01.010]
19. Topol E. Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again. New York City, New York: Basic Books; 2019.
20. Contreras I, Vehi J. Artificial Intelligence for Diabetes Management and Decision Support: Literature Review. J Med Internet Res 2018 May 30;20(5):e10775 [FREE Full text] [doi: 10.2196/10775] [Medline: 29848472]

21. Zellweger MJ, Tsirkin A, Vasilchenko V, Failer M, Dressel A, Kleber ME, et al. A new non-invasive diagnostic tool in coronary artery disease: artificial intelligence as an essential element of predictive, preventive, and personalized medicine. EPMA J 2018 Sep 16;9(3):235-247 [FREE Full text] [doi: 10.1007/s13167-018-0142-x] [Medline: 30174760]
22. Shandler R. The age of surveillance capitalism: the fight for a human future at the new frontier of power. Journal of Cyber Policy 2019 Jul 07;4(2):303-304. [doi: 10.1080/23738871.2019.1637914]
23. Safavi K, Mathews SC, Bates DW, Dorsey ER, Cohen AB. Top-Funded Digital Health Companies And Their Impact On High-Burden, High-Cost Conditions. Health Aff (Millwood) 2019 Jan;38(1):115-123. [doi: 10.1377/hlthaff.2018.05081] [Medline: 30615535]
24. Byambasuren O, Sanders S, Beller E, Glasziou P. Prescribable mHealth apps identified from an overview of systematic reviews. Digit Med 2018 May 9;1(1):12.
25. Wollowski M, McDonald J. The Web of Smart Entities-Aspects of a Theory of the Next Generation of the Internet. Artificial Intelligence for the Internet of Everything 2019:117-137. [doi: 10.1016/b978-0-12-817636-8.00007-7]
26. Morley J, Floridi L. The Limits of Empowerment: How to Reframe the Role of mHealth Tools in the Healthcare Ecosystem. Sci Eng Ethics 2019 Jun 06:1-25. [doi: 10.1007/s11948-019-00115-1] [Medline: 31172424]
27. Fiske A, Henningsen P, Buyx A. Your Robot Therapist Will See You Now: Ethical Implications of Embodied Artificial Intelligence in Psychiatry, Psychology, and Psychotherapy. J Med Internet Res 2019 May 09;21(5):e13216 [FREE Full text] [doi: 10.2196/13216] [Medline: 31094356]
28. Kickbusch I. Health promotion 4.0. Health Promot Int 2019 Apr 01;34(2):179-181. [doi: 10.1093/heapro/daz022] [Medline: 30939200]
29. Topol E. The Creative Destruction Of Medicine: How The Digital Revolution Will Create Better Health Care. New York City, New York: Basic Books; 2019.

Abbreviations

AI: artificial intelligence
eHealth: electronic health
EHR: electronic health record
HIPAA: Health Insurance Portability and Accountability Act
IoT: Internet of Things
JMIR: Journal of Medical Internet Research
JoPM: Journal of Participatory Medicine
WHO: World Health Organization

Edited by G Eysenbach; submitted 23.09.19; peer-reviewed by S Kitsiou, Y Zhou, D Goyal; comments to author 14.10.19; revised version received 23.10.19; accepted 11.11.19; published 21.11.19.

Please cite as:
Kukafka R
Digital Health Consumers on the Road to the Future
J Med Internet Res 2019;21(11):e16359
URL: https://www.jmir.org/2019/11/e16359
doi:10.2196/16359
PMID:31750835

Discussion Paper

Gimme My Damn Data (and Let Patients Help!): The #GimmeMyDamnData Manifesto

Dave deBronkart[1], SB; Gunther Eysenbach[2], MD, MPH, FACMI

[1]Society for Participatory Medicine, Nashua, NH, United States

[2]JMIR Publications, Toronto, ON, Canada

Corresponding Author:
Gunther Eysenbach, MD, MPH, FACMI
JMIR Publications
130 Queens Quay E, Ste. 1100
Toronto, ON
Canada
Phone: 1 416 583 2040
Email: editor@jmir.org

Abstract

Ten years ago, in 2009, “e-Patient Dave” deBronkart delivered an influential keynote speech at the Medicine 2.0 conference in Toronto, organized by the *Journal of Medical Internet Research*’s (JMIR’s) editor-in-chief Gunther Eysenbach, who themed the conference around the topics of participation, openness, collaboration, apomediation, and social networking to improve health care for the 21st century—with patient participation being a major component. Many see this as a defining event within the participatory medicine movement, perhaps the beginning of a social movement, similar to the women’s rights movement, with the title of Dave’s keynote “Gimme my damn data” becoming a rallying cry and hashtag for patients demanding more access to their electronic health records. On the occasion of the 20th anniversary of JMIR (and 10 years after the keynote), we are celebrating the impact of the keynote for the participatory medicine movement and #gimmemydamndata (also #GMDD) by publishing the transcript of these initial conversations as a manifesto of patients’ rights to access their data and their right to save their lives.

(J Med Internet Res 2019;21(11):e17045) doi:10.2196/17045

KEYWORDS

data; participatory medicine; ehealth

Introduction by Gunther Eysenbach

Ladies and Gentlemen,

It is my great pleasure to welcome you to the second Medicine 2.0 Congress, also known as the World Congress on Social Networking and Web 2.0 Applications in Medicine, Healthcare, and Biomedical Research.

One of the most difficult tasks of being a chair of this meeting is, perhaps, that each year, I see it as one of my obligations to define or at least outline what we mean by Medicine 2.0 or Health 2.0, which seems to be getting increasingly difficult. Last time I checked, I found about three dozen different definitions of what people think Health 2.0 or Medicine 2.0 is. However, I hope that at the end of the two days, you come out of this meeting with sort of a warm, fuzzy feeling and you have a feel for what it is when we talk about Medicine 2.0. As a side note, I should maybe stress that I see the terms Health 2.0 and Medicine 2.0 as being largely interchangeable.

Health Care Needs to Change

Despite this problem of lacking a clear definition, I think what brought us all together here—the kind of overarching theme of Medicine 2.0—is the recognition that health care needs change. No matter what country we are coming from—and by the way, there are people from 23 different countries here in this room—no matter what health care benefits we enjoy, no matter what health care system we are “enjoying” (or suffering from), no matter how we fund health care and what percentage of gross domestic product we invest in health care, what we all have in common as patients, consumers, and health care professionals is that we are becoming increasingly frustrated with the way health care is delivered, frustrated and impatient with medicine and health care, which often uses antiquated processes and technologies. Maybe one reason for this is that we are all part of the Google generation, and we are used to getting data, information, and answers at the click of a button, which seems to work in almost every other industry—except health care. Of course, there are reasons for that: privacy first and foremost. However, it is incredibly frustrating for us as consumers,

patients, and health care professionals, especially for those of us who know that the barriers are not of a technological nature. There are so many things, which, from a technological point of view, are feasible. However, the barriers are mostly of an organizational nature, cultural nature, or political nature, and there is an increasing gap between what is technologically possible and what the status quo is. So, one possible way of looking at or defining Medicine 2.0 is to see it as "next-generation medicine": what medicine and health care could be and what it should be, and hopefully, what it will be in the near future—if we all work together, if we are smart, if we reengineer health care using technology and participatory approaches, and if we reengineer health care using Web 2.0 values and approaches.

Participation and Other Medicine 2.0 Values

Now, what do I mean by "Medicine 2.0 values"? I wrote a paper about Medicine 2.0 [1], where I outlined what, in my mind, the five characteristics of Medicine 2.0 or Health 2.0 are: (1) participation, (2) openness, (3) collaboration, (4) apomediation, and (5) social networking.

I think these five values or characteristics of Medicine 2.0 are the perfect remedies for what health care and medicine are suffering from in many countries, especially industrialized countries, worldwide (Figure 1).

These problems include, for example, a focus on curative medicine; insufficient incentives for prevention; and insufficient incentives to keep people out of the health care system, keep them healthy, and keep them out of hospitals. To keep people healthy, we need participation, specifically end-user participation. We need the participation, involvement, and engagement of consumers and patients, and that's one of the values of Medicine 2.0: *participation*.

Health care around the world suffers in many cases from intransparencies and hierarchies, and if you look at computer systems, they are mostly proprietary systems, closed systems. Again, the Medicine 2.0 value of *openness*, sharing data, sharing experiences, and sharing outcomes provides a possible remedy to that. Health care is often characterized by information silos—inadequate patient access to information—and again, regarding the Medicine 2.0 value of *collaboration*, interoperability—the idea of patients as partners—provides an antidote to that.

In health care, we often deal with intermediaries, information brokers, and gatekeepers. On the other side of the equation is the Medicine 2.0 idea of wisdom of the crowds bypassing intermediaries, gatekeepers, and experts—not replacing them, but complementing them with wisdom of the crowds.

Lastly, social networking. Especially as medical informaticians, it is funny that we have focused our efforts on modeling medical information, and not on modeling and storing *relationships* between people. That is what social networking is all about.

Figure 1. Current health care problems vs Medicine 2.0 (Eysenbach).

In social networking, what we are trying to do is storing and modeling relationships between people, and that adds an entirely new dimension to what we deem as health care information. Social networks are fascinating because they also allow us to

enable and facilitate collaboration between people. For example, they allow for collaborative filtering processes, enabling identification of relevant and trustworthy information based on what peers are doing, and for reputation and trust management. It allows viral dissemination of information and apps, which is a possibility that especially excites public health professionals, and you are going to hear some examples of that at this conference, for example, the use of Facebook to disseminate information. It is a potentially powerful tool to engage end users and patients in health behavior change apps and similar apps, because the common problem many of these fantastic Web-based behavior change tools suffer from is what I called the "law of attrition," of people starting with a high degree of enthusiasm using these kinds of systems, but then they drop out. They discontinue the use of these kinds of systems. I think that if you add virtual communities or social networking, if you add some sort of peer pressure and social support to people to return to the website—not only peer pressure but some additional incentives, for example—they find their friends or peers there, that is a very powerful tool to engage patients.

So, some of the questions this conference hopes to answer or at least to raise are as follows:

- What are the implications of all these technologies for health and health care policy?
- How do the generic Web 2.0 concepts and technologies translate into health apps?
- What are the specific requirements for health-related or medical social networking apps?
- What are the research questions and issues and what are the methods to answer these questions?
- What are the determinants of success or failure?
- Is the "hype" promoted by private enterprises and venture capitalists supported by evidence and what can we expect for the future?

Introducing "E-patient Dave"

It is now my great pleasure to introduce our keynote speaker, e-Patient (electronic patient) Dave deBronkart.

About 32 months ago, Dave was diagnosed with Stage 4, Grade 4 kidney cancer. It was bad, with metastasis to the lungs, several bones, and tongue. He read that his immediate survival time was 24 weeks. Using the internet in every way he could, he learned everything he could about his treatment options, joined an expert patient group on ACOR (Association of Cancer Online Resources), built a social network, and shared online medical records with anyone who could help. He beat the odds.

Having faced death, 2 years ago, he asked himself what he would do with his free replay in life and became a blogger about health care. Then, his doctor, Danny Sands, invited him to join the e-Patient movement, where he has become one of the most outspoken advocates for patient empowerment and patient engagement. As some of us who follow his blog know, he is a data guy. This March, when he clicked on the button to move his data from his hospital to Google Health, the mess that resulted was front-page news in the Boston Globe. He first blogged about it, and then, the media picked it up. That was around the time when I invited him to be a keynote speaker at Medicine 2.0.

His insightful blog post about the problems in interoperability and the publicity that followed led to him being invited to a lot of policy meetings in Washington. We are proud that we are the first conference where Dave is a keynote speaker, and I think it is not going to be the last one. Last month, he launched a consulting practice and just this week, he is featured with a full-page photo and a cover story entitled, "The Patient of the Future," in Health Leaders magazine

Ladies and gentlemen, please join me in welcoming the "Patient of the Future": e-Patient Dave (Figure 2).

Figure 2. E-patient Dave deBronkart and Gunther Eysenbach (Founder and Chair) at Medicine 2.0 in Toronto in 2009.

E-patient Dave deBronkart

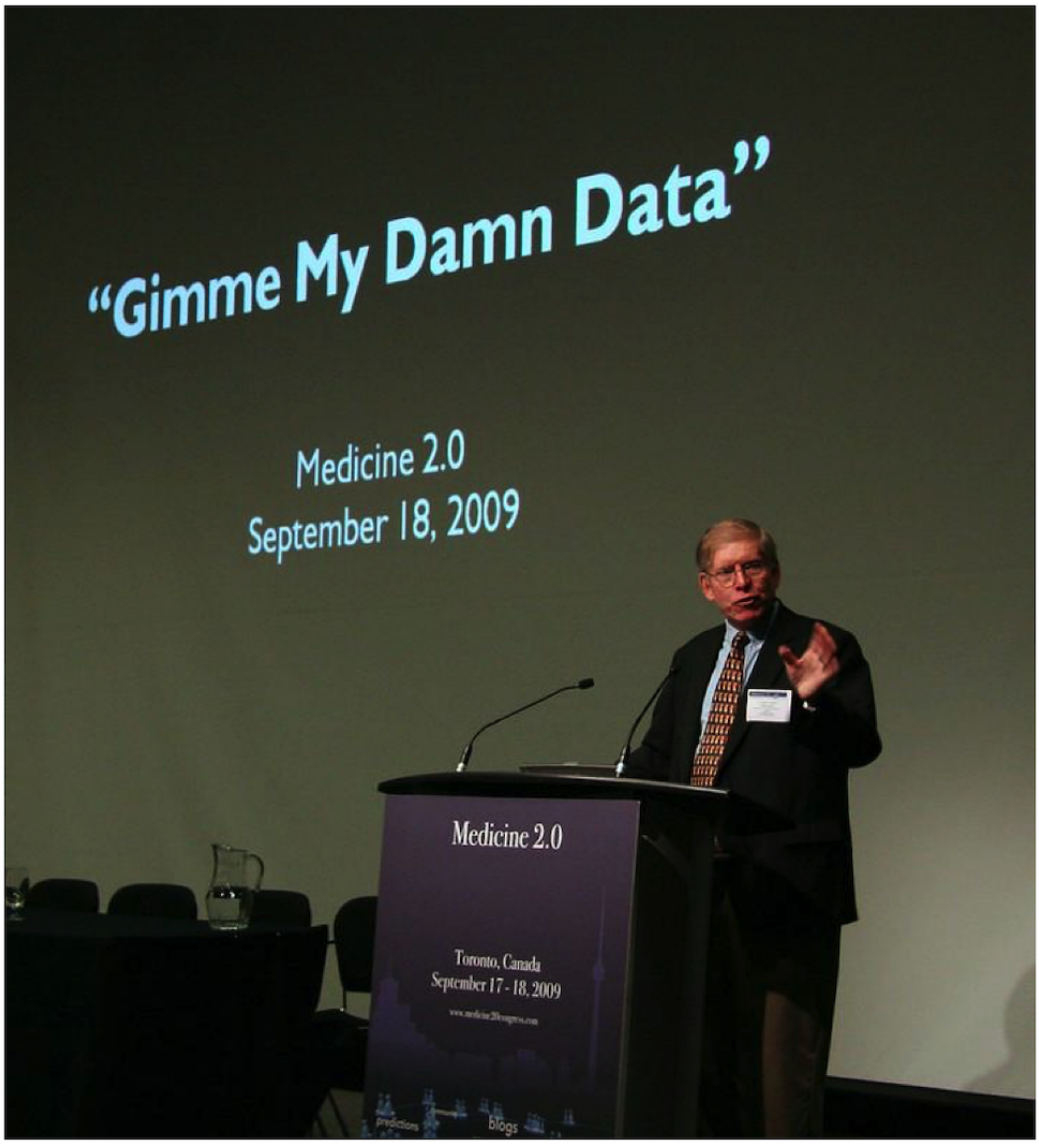
Introduction

What a thrill! Thank you so much! Gunther (Eysenbach), thank you for your vision. People who do not know the health care world are stunned when I tell them that a patient is giving the opening address at a conference like this. For those of you who do not know as much of Gunther's history, he is one of the original visionaries who did the initial bit of research that, in a sense, opened up the whole world of participatory medicine: A number of years ago, people were concerned about the question of the danger of going out and googling, because we all know there is garbage out there. He researched it (among other things), and for years, tried to find a single case of death by googling, and he did not find it. He published that result. He even added a bounty after a few years of failure. That is a seminal moment, a vital piece of information in Tom Fergusons' e-Patient paper, that I will talk about briefly later.

So, the title of this talk is indeed, "Gimme My Damn Data!" (Figure 3). The subtitle was going to be "Because You Can't Be Trusted With It." However, I toned it down a little because this is being recorded and broadcasted.

Talking about participatory medicine, the nice-looking guy on this slide, somebody thought that was clip art. No, that's Tom Ferguson, the guy who founded the e-Patient movement and became an e-Patient himself in his later years. His movement has now become *The Society for Participatory Medicine*, which has just this year launched ParticipatoryMedicine.org. We are not really active on Twitter yet, but the address will be @S4PM. Aren't you glad we did not spell out @participatorymedicine? I just hate long Twitter names.

Figure 3. "E-patient Dave" deBronkart's Opening Keynote at the Medicine 2.0 conference. [Editorial Note: The slide shows the wrong date; the actual date was September 17, 2009]



Foundation Principles

Here are some foundation principles:

1. Patient is not a third-person word. Your time will come
2. The right of a desperate person to save themselves
3. The right to know what your options are
4. The right to pick up your data and pursue an option elsewhere

I take this very seriously. Gunther told you a little bit about my story. However, my one story does not change the world. As I have become involved in this, I found myself, as Gunther said, asking myself, "What am I going to do with this free replay in life?" I found this calling to health care and as I started studying it. I have only been at this for a year and a half; the first session I went to was a small high-tech conference outside Boston. When my turn came to talk I said, "The thing I want us to shift here is that patient is not a third-person word, alright? We talk about patients as if it's somebody that's not here in the room right now. Well, trust me your time will come, alright?" and I went on at some length about that.

The next thing that came up: I went to a privacy meeting in Washington discussing several issues: the HIPAA (Health Insurance Portability and Accountability Act), data privacy, and others. We have so many obstructive regulations in place that it interferes with the right of a person to try to save themselves. Privacy and data security take on a whole different meaning when your life is at stake. I mean, look at it this way: We take our pants off for the doctor—why would we not share information to save our lives?

Then, "the right to know what your options are": This is important with social networking (which helps to find these alternative options).

Ultimately, when necessary, the right to pick up your data and pursue an option elsewhere: "You know what? This isn't working here. I want to go to this other hospital. Give me my data."

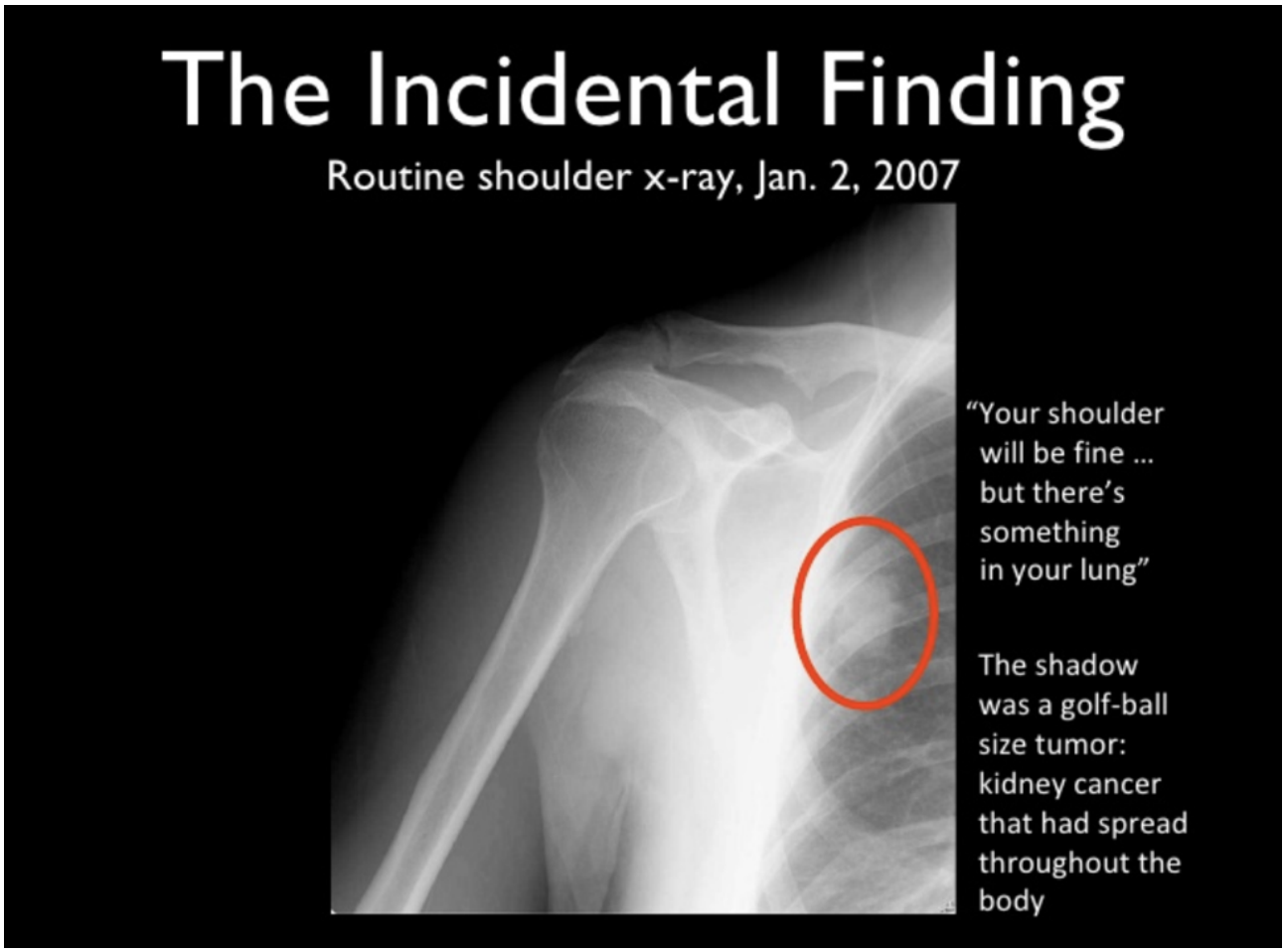
My Story and the Realities That We Deal With

Technological innovation can vastly alter things—iPods, cell phones, computers. However, health care is, in many ways, far behind other industries, and yet, the good news is it is not rocket science. The tools are available. Technologies are available. We just need to start doing this. Honestly, I am becoming impatient with the slow rate of change.

Here is a quick review of my story:

I had a sore shoulder at the time of my physical 3 years ago, and I got a shoulder x-ray. Now, those of you who know how to look at x-rays will see that there is something there, that shadow should not be there (Figure 4). Totally by coincidence, the shoulder x-ray picked up a golf ball–size lesion in my lung that happened to be near the shoulder. That shoulder x-ray saved my life because I did not have any physical symptoms until 6 weeks later and by then, it would have been too late.

Figure 4. Dave's x-Ray.



So, that was ultimately found to be kidney cancer that had spread throughout my body. I was near the end. So, I researched (I have always been a Googler) as I like to say, I googled my butt off. This happens to be a graphic from the website of the treatment that I got, which was high-dosage interleukin.

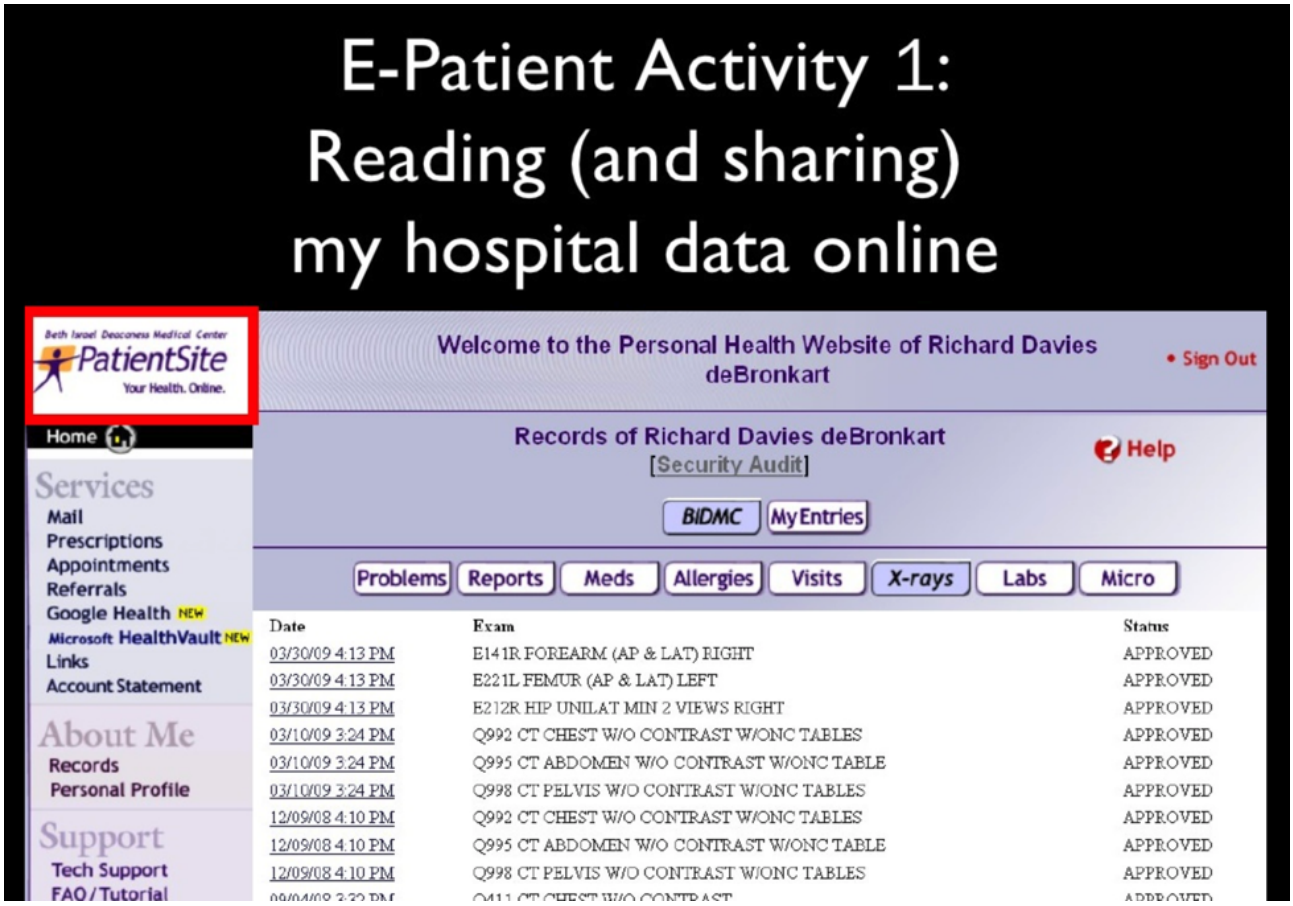
Indeed, in every one of those spots, I had a tumor. I had one in my skull, numerous throughout my lungs. The one in the femur was so big that eventually, the leg broke. I now have a nice steel leg. That is a pretty big metastasis when it breaks your leg. Here are the others. There was one in my ulna. When the leg was in the process of breaking, I could not use a regular walker because it put too much stress on the ulna, which also had metastases in it. Finally, one in my thigh muscle and in my tongue. That is gross. I am glad that happened just before the treatment started because it pretty quickly would have fallen off.

It is sobering. I mean, I have always been somebody who has been able to find ways around things. Whether it is computer software or car research, I do not do well with being told, "You can't do that." However, every place I looked, I read things like, "The prognosis is poor. Almost all patients with Stage 4 Renal cell cancer are incurable." I remember the night that the biopsy finally confirmed the diagnosis, and I read "24 weeks survival time." Now, this was in January. I remember waking up at 1 AM one morning and thinking, "I might not see Christmas. I probably will not see my daughter's wedding." I actually had the image of seeing my mother's face at my funeral, and I faced the task of sitting with my daughter and her boyfriend and giving them the news and telling them that they had better not get married prematurely just so they could do it while I was alive.

So, you are left with several questions: "What are my options?" "What can I do?" "How can I get myself into gear?" So, you get engaged. You do everything you can. First thing I did was go to PatientSite, my hospital's personal health record system (Figure 5). It is ugly and really needs a makeover. However, you know what? I could look at my data, and I could give my

password to other people so that they could look at it and give me advice and feedback as well.

Figure 5. Dave's personal health website at Beth Israel Deaconess Medical Center.



Patient Communities

The second thing I did, I joined ACOR (Association of Cancer Online Resources). I love this phrase: "My doctor prescribed ACOR" (Figure 6). My doctor handed me a prescription slip with "ACOR.org" written on it [Editor's note: This is now SmartPatients.com], and I found there was more useful, action-worthy information from other patients than I found on any encyclopedia-style website. Encyclopedia-style websites could give me peer-reviewed information that was 10 years out of date and still could not tell me anything about what the treatments were really going to be like.

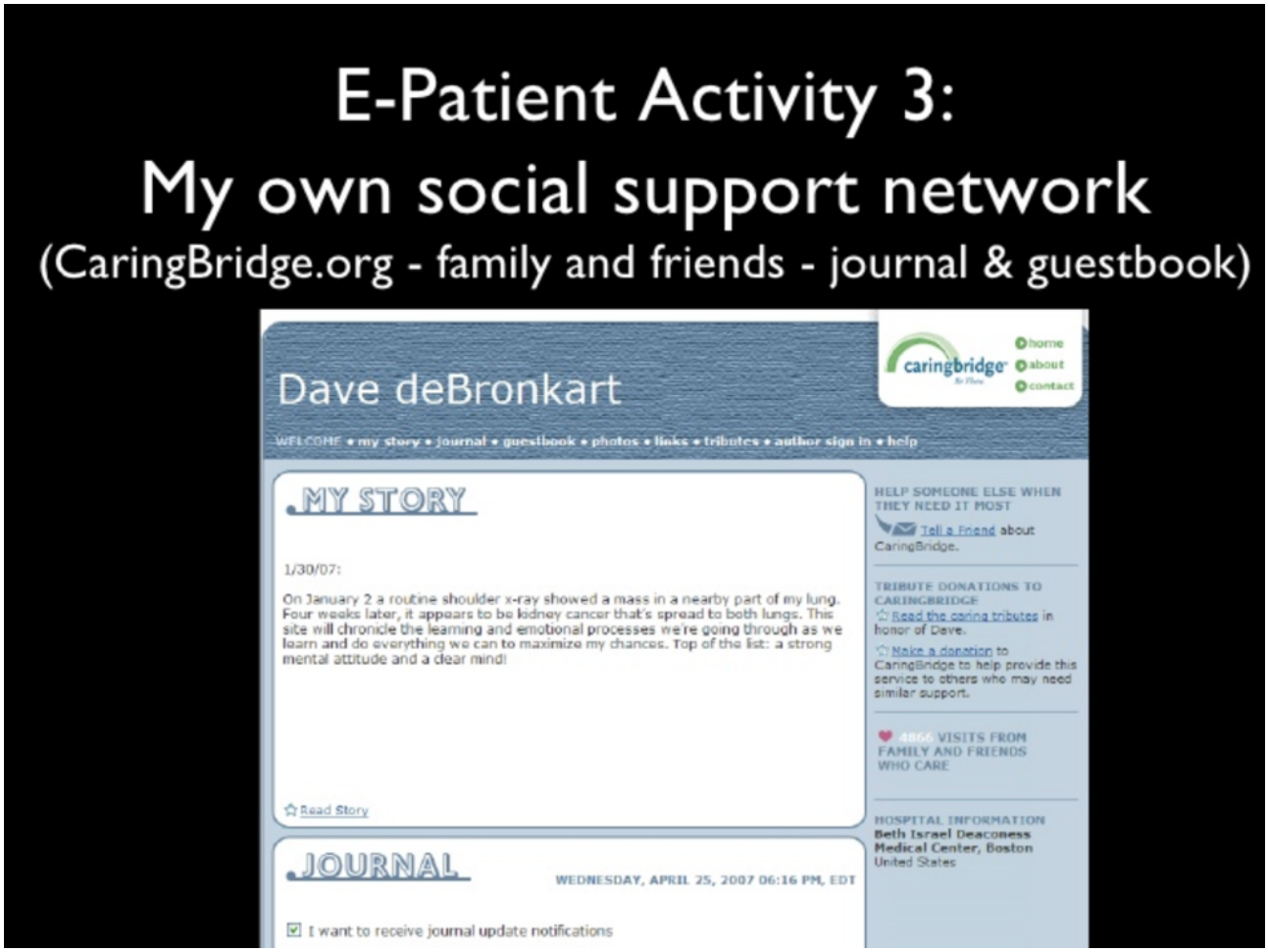
Patient communities are responsive. People discuss what to do, and patients know what patients want to talk about. Consider this, we talk about referral delays reaching a doctor. Well, I got responses in 4 minutes and 11 minutes to my first questions in the patient community.

This is an appeal I always toss out: Whatever we do in health care spending, we should just devote 1% of it to funding patient communities, to let patients do what they see as necessary. For those of us who think about how hard it is to get physicians to adopt medical record systems, consider this: Down, at the bottom, is an audience who you do not have to motivate for adoption. These people are already motivated and engaged.

Finally, my own social support network (Figure 7).

Figure 6. The Association of Cancer Online Resources' cancer patient communities.

Figure 7. Dave's own social support network.



Participatory Medicine

Now, long story short, because I have other things I want to get onto, the treatment worked. Surprise, surprise! I made sure I got copies of all my scan images. There was one of the lesions on the left side before. If you consider the size of my midsection, that is was a pretty good-size tumor. That little white dot is what was left of it 50 weeks later. The tumors have continued to shrink, and the ones that remain have been stable and are considered to be dead. So basically, I won. So, what do I do now?

In January of 2008, as Gunther said, my physician Danny Sands invited me to join this group he belongs to, "Equipped, Engaged, Empowered, and Enabled: E-patients - A Different Kind of Beast." There is a white paper by Tom Ferguson, that, if you have not read, this is the paper that mentions Gunther's contribution as well. It is at the top of the e-Patients.net website. One key point was, and this was a mind pop for me, that Donald Lindberg, the head of the National Library of Medicine said, "If I read two journal articles every night, at the end of a year I'd be 400 years behind." What was a mind pop about this for me was, "Okay, I can help." It is not a case of the doctors knowing everything they need to know. I can help.

Then, there is the other important point: the lethal lag time.

The publication delay between the time information comes into reality and the time it takes to make its way through the publication pipeline is 2-5 years. In fact, from what I have heard, the time from when an idea is conceived until the research is conducted and the results are out there can be 17 years. During that time, people can die due to a lack of information that exists. The internet can solve this. Tom Ferguson had this early vision of how accessed information would turn health care on its head. These slides are on e-Patients.net too. The idea here, with industrial-age medicine before the internet, was, if you look at the bottom, that the ability to create value, like in the 1800s, accrued to people who owned a factory. That was the means of production. However, the information-age medicine turned health care on its head and now, what we have are individual self-care friends and family. The latest Pew Internet research data says that people turn first to friends and family and then to self-help networks like ACOR and then to professionals as facilitators. I will come back to this point of professionals as facilitators.

The key is that the internet gives us access to each other and to information that, in the past, we could not get at—and that changes everything when you are desperate. That is participatory medicine. Tom Ferguson saw this within a year of birth of the Mozilla browser. That was an amazing insight. He published those slides in January 1995.

Regarding this point about participatory medicine, I wrote a post in December about Stanley Feld, a physician who had written a post called, "Physicians Are Coaches and Patients Are

Players." I will not spend time on it here, but think about this. In reality, physicians advise us on what to do, but we do not always do it. For example, I take my medication, but exercise and eat better? Well, not so much. What you start to see, when you give up the viewpoint, is that although they have all the knowledge and power, "a lot is up to me as to whether I do what they say." Things start to shift. More about that later.

Decentralization Follows Centralization

I work in high-tech software marketing. I have been involved in high-tech industries around Boston for my whole career. This spring, I met Clayton Christensen, the author of *The Innovator's Prescription*, and all of the innovator's dilemma books. He has studied this for years and has come up with a really world-changing perspective on how new technology changes an industry. His health care book was published in February. There are a lot of people who disagree with it. I do not assert that everything he says is the solution, but I want you to think about something that I know, from my personal experience, is really accurate.

In the bottom left is an archaic device called a slide rule. I used to own one, and one of Christensen's points is that the centralization of a skill or a technology is followed by decentralization, and here is how that played out in computing:

We have here a picture of an old-fashioned computer room, the kind of thing where, when I was in college, we would take a stack of Hollerith cards to the computer room and hand it in and come back the next day to get our printouts: This involved highly specialized skills, expensive equipment, and specialized environment. As time went by, it got to the point where things were more understood and you could have a simpler, smaller computer in a less specialized area. So, you had departmental mini-computers. Then, it got to the point where we had desktop computers, and then, ultimately, laptops, and now, today, it is on our hip with a smartphone. These innovations depend on the processes becoming reliable, so you no longer need a computer department geek *[sic]* to run the machine, and as the data is mobile, it can move around. The decentralization that follows centralization is only beginning in health care, but it is beginning.

Here are three pictures of a family doctor, like we had when I was growing up, who would come to our house and in the center now, like a computer room, we have the massive general hospital. I met Christensen at a meeting at Beth Israel Deaconess in Boston. It was a promotional event they were doing to encourage people to fund research there. Christensen stood there and looked Paul Levy, the Chief Executive Officer, in the eye and said, "The general hospital is not a sustainable business model. It cannot survive without philanthropy and government subsidies." That is a pretty challenging thing to think about. Why? I will get to that in a moment.

So, here you have, as with computer departments, all kinds of specialized equipment and the general hospital is set up to be able to do anything for anybody. What has happened now is that there are smaller regional medical centers that can do a lot of things that used to require going downtown or to the big centers. Now, there is more technology in the physician's offices, and finally more things are being done at home with connected health equipment and self-monitoring, and this is only the beginning.

Here is a diagram that Christensen used for this—a metaphor of a general hospital. This is a diagram of a hypothetical factory, a machine tool department that can do just about anything for anyone. So, with all these work stations, a particular product might start at one station and wander all over creation. This is the path taken by product A through this shop. This is what it is like when a patient goes through a general hospital with a complex condition. Here is a different product that goes through a completely different path.

Regarding this, I met with Jason Wong, the co-author of the book and said, "Do I understand this correctly? Alright, let's say instead of that being an assembly line, let's say this is someone's life. So over on the left, you have 'Life starts here' instead of 'Product A starts here' and 'this is the path taken of life.'" This starburst here, this is the moment of my diagnosis, the moment of awareness. Everything in Christensen's and Wong's book is about the delivery of care after the diagnosis when it is recognized that something needs to be done. What is not mentioned there is that today with new technology and social networks, values being created outside the hospital walls and outside the health care delivery systems, what if this awareness happened earlier in the process? What is the impact of having the information earlier? Well, in my particular case, as a kind of a trivial example, I got diagnosed 6 weeks earlier, so I am alive. That is an impact that I care about. On the other hand, there are other ways that we could do better and be more aware of our well-being. Making a difference back at this point requires two things:

- Data: The evidence of what is happening
- Knowledge: The meaning of the data and what to do about it

That is the meaning of an engaged, empowered patient with participating providers. This is how it goes differently. You are more aware if you have the information and the knowledge of what to do about it.

Google Earth for My Body

Here is another example of this—and Gunther talked about technologies that are out there—and we are starting to use them to match them up in ways that have not been anticipated.

One of the tumors in my kidney was encroaching on the psoas muscle. Now, I had no idea what a psoas muscle was. I went to visiblebody.com, which is this amazing, free, interactive, 3D website. I cannot believe they can do this for free: You can turn on the skeletal system, the nervous system, and so on. You can click on things to remove them, and I was able to create this image of the relationship between my kidney and the psoas muscle and then rotate it in 3D. Look through. I thought, "What the heck!" We have Google Earth where you can fly to Duluth and go to an individual building; why not Google Earth for my body? This is a generic body, but it is not hard to imagine that we take my scan image, which is currently a crummy, black-and-white, low-resolution thing and mash that up with this, so that we can actually look at my body in 3D. I can tell

you from a visceral level, I have a different relationship to what is going on inside me, given this awareness of where the pieces are.

This is the article "Patient of the Future" from this week's *Health Leaders* magazine that Gunther mentioned. You can see this online. I've got it here because I just love the vision that the woman, Gienna Shaw, articulated. She was new to this whole subject, but she completely got it:

> *Patient enters the waiting room and is greeted by her personal navigator who hands her a tablet size computer. In his office, the physician is reading an email from a patient who has forwarded an interesting study.*

There was a great moment almost two years ago where I was waiting for a CAT scan follow-up visit in the doctor's office. While I was waiting, I was on the computer doing some stuff, the oncologist and the nurse practitioners came in, and a few minutes later, my wife cracked up. She was sitting there. She said, "What's wrong with this picture?" because I was sitting there pointing something out on the computer while the oncologist sat on the examining table and the nurse practitioners were writing down the URL. It is a new world, people, completely new world.

Where Are We Today? The Google Health Disaster

So, that is the vision. Where are we today? I went to move my data into an online personal health record. So, I punched the button to move my data into Google Health, and I got this craziness. I got this false medication warning saying that my blood pressure medication conflicted with low potassium in my blood. Well, low potassium in my blood was true when I was in the hospital being treated for the cancer. It was not accurate. Plus, there was a whole bunch of things: In my condition list, on the right side, they listed everything I had ever had with no dates attached to it, which was crazy. We looked into it, and it turns out that they had transmitted billing codes instead of clinical data, and, from an IT (information technology) perspective, it is just goofy to pick up one type of data just because it happens to be available and use it even though it is not appropriately modeled for what you are going to do with it.

Here is the thing: It is a long story. You can read the 3300-word blog post about it on e-Patients.net if you want. I have to say that both the hospital and Google Health responded in an exemplary way. They completely disclosed everything. Well, once it hit the front page of the Boston Globe, they responded in an exemplary way. However, I got a complete Excel spreadsheet of every billing record that the hospital had for me, and I went through and added notes on the left. My physician went through and added notes on the right. We had some crazy things like the top note on the right points to an item 424.2 "non-rheumatoid tricuspid valve disease." Well, for one thing, I never had that, and for another, that was noted during a visit where I was getting an infusion to treat my bone metastases. It is not normal for a cardiac condition to be diagnosed during an orthopedic visit.

Down at the bottom, bone and cartilage disease. That was during the visit where I went in when the tumor had erupted from my tongue. How that billing code got in there, nobody knows. These are not data that are well managed, and this is at a hospital, by the way, that is well known for being one of the leading hospitals in handling IT.

Another example is the top right "Aortic Aneurysm." Well, it turns out that on one report, a radiology report showed a 1/4-inch enlargement at the base of the aorta, and strictly speaking, that is an aneurysm. Do you know what the trick is? Upcoding. You can bill more if an aneurysm is involved than if there is a slight swelling involved. When I learned about upcoding, I immediately thought about my supermarket, wherein the deli section, they have regular bologna and then they have, next to it, something that looks awfully similar, labeled "Tasty, delicious bologna" and it is priced higher. So, here we have tasty delicious bologna applied to our health care cost.

Then, there are other things like volvulus of the intestine. This is a life-threatening kink in the intestine, which is fatal if not fixed in a couple of days. I have never had it. How it got in, I do not know. How much of the excess health care cost in the United States is due to things like this, and nobody can track it down. So, physician errors, clerical errors, and upcoding.

Here is the thing, HIT (health IT) just needs to follow normal IT best practices. Find the right data vocabulary. Use good reliability practices and test them with real-world data before going live. It is not rocket science; it is available to us.

Here is another risk that nobody talks about. In the United States, there is an organization called the MIB (Medical Information Bank): It is an insurance industry association where insurance companies share what things you have had billing codes submitted for, so that you cannot hide conditions from them such that they might be harmed by your dishonesty. However, this video from Consumer Reports is about a woman from the Katrina area in Mississippi who lost her 401K savings and everything else. She has asthma. Her insurance company would not cover her treatments, and it turns out the root cause was that her physician had accidentally miscoded something and it took her forever to find that out.

The MIB is a private insurance industry data bank. I got in touch with them. They are totally opaque. You cannot find out stuff in there. Because they are similar to a credit bureau, you are allowed to get a copy of your records, but as I wrote on the blog, and I would not go into it here, but good luck making sense out of it. Their lawyers started writing to me and made clear that they do not consider themselves liable for any damage they cause. So, the lesson? You better check for errors. If you are an American, check into the MIB.

Here is the question, what is in your wallet medically? Do you personally know everything that is in your medical record? HealthDataRights.org published the declaration of health data rights this summer. You can go sign it, endorse it if you want to. At Health Camp Toronto, yesterday, we had a terrific discussion initiated by Jen McCabe who will be on later this morning saying that this is not tough enough, that one thing that is not stated here is the fundamental thing: This is my data. It is my property. I mean, whose data is it? So, we are actually thinking about revising that. My endorsement said, "Look, these

rights are as inalienable as the right to life itself. Whose life depends on its accuracy? Whose data is it anyway?" My physician in his endorsement said, "How can patients participate if they can't see the same data?"

To wrap up, Clay Shirky is a real internet visionary. For instance, if you are familiar with the concept of the long tail, which really alters our thinking about finding things on the internet, he is one of the developers of that idea. On the subject of data sharing, he said, "Giving patients access to their medical records will naturally improve the quality of what's in there." I love the way he said it. You clean up when you know company is coming, when you know somebody is going to be looking at the data. He also pointed out the opposite model—clean then share never works out because it turns into, "Well, we can't share it yet because we haven't cleaned it yet," and in the meanwhile, lives are at stake.

Finally, Gunther talked about Medicine 1.0 and Medicine 2.0. We had Web 1.0, which was publish-only. Web 2.0 is read-write. We can put things out on the Web. We can share information with each other. No less an authority than Tim Berners-Lee, considered the founder of the Web, gave a talk at TED Talks this spring where he talked about what is next for the internet, and it's beyond anything I had realized. It is not just aggrading similar pieces of published data. He is talking about intelligent agent software that can go out and derive new information, create new knowledge. However, it cannot do it if it is looking at other people's interpretations. It has to be looking at the raw data. So, his "TED Talk" is a great video. It ends with him actually getting people to start chanting, "Raw data now! Raw data now!" We, I think, ought to be thinking along similar lines.

Takeaways: It is my data. It is your data. Innovation depends on good-quality data and reliable workflows. Let us help.

Let Us Get Involved—Give Us Our Data!

I got to the point this year, where I said I cannot keep doing the day job and do this at night. So, this is my disclosure of financial interest. I have a financial interest in keeping myself funded so that I can do this. I have gone into business for myself.

To wrap it all up, here are 2.8 years in pictures. This is me 3 days before that diagnostic x-ray. I did not realize it at the time, but I was not looking too good. Then, by Halloween, in October, at the company Halloween party, this was me kind of grinning at the reaper. I did not lose my hair due to chemotherapy. Kidney cancer does not use chemotherapy. However, my hair became an absolute mess, and I just said screw it and chopped it all off. So, that is me very happy 10 months later. This is me with my mom, not at my funeral, but at my daughter's wedding in May. It was a great thing. It was wonderful. Here is a real kicker for you. This is me last Sunday with my bone surgeon, because somehow, something has happened to me and I have decided to get interested in my health, and I rode a bike in a cancer fundraiser. I have not ridden a bike since I was 20 years old. However, this June, I said to my wife, "You know what? I want a bike." We went and got a yard sale bike, and it turned out that I loved it and I was good at it. Just to show that this health care stuff is really serious, I lifted up my shorts, got out a magic marker, and wrote a note to my bone surgeon, "cut here," to avoid any chance of any wrong side errors.

That is it. Here is something to think about. Could it be that engaged patients might take better care of themselves? Once you really get that all the knowledge that is not out there, that I have something to do with it, that is my coach. I am the one on the court playing the game. So, let us get involved. Give us our data!

Thank you.

Q&A With Introduction by Gunther Eysenbach

Thanks very much, Dave. You almost got people chanting, "Give us our data!"

We have about 25 minutes for discussion and questions. In general, we try to leave ample time for discussion at this meeting because we want this to be an interactive meeting. There are two microphones here in the auditorium coming around. First question:

Audience Member #1: "It is a Social Question"

Dave, I mean, you are absolutely correct. I applaud you on a really excellent presentation, and congratulations on beating your cancer. You are a very strong individual and someone to be looked up to.

I think that you have hit upon one of the most essential issues in health care today, and I can tell you that as a physician, I remember treating patients, and I have treated patients, but I remember one, in particular, that came to me with lesions in the brain and needed a tumor resected. I had actually already received this information and someone suggested just watching it. We removed the tumor and found out it was a melanoma. After surgery, she came back and we were talking. The question came up, "Have you ever been diagnosed with melanoma before?" The answer was, "No. Never had a problem." She left the office, I was seeing the next patient, and she came back in. She popped her head in the window and said, "You know, I did have a skin biopsy about two years ago. I assumed it was normal. I never heard anything about it." I actually knew the dermatologist. He was in the same building that I was working in. I called him, excellent reputation, excellent dermatologist. I mean, renowned in the area, and I asked him to look up the biopsy results. He said, "Yeah sure, Luke, you know. I'll be right back." He came and there was a moment of silence. I heard him rustling papers. After a moment of silence, he says, "I don't know how this happened, but the result was in the chart. It was melanoma. There's no signature. I sign and date every paper I've seen. Someone in my office popped it into the paper chart without my having a chance to review it." The patient was upset, of course. However, she accepted it as human error, and she passed on about 7 months later.

Since then and with other episodes, I see this happen over and over again. The issue is essentially that: How do you get the data to the patient? It is really not even a technology question. *It is a social question*. Just getting your own data, even as a

physician, out of a hospital is almost impossible. Sometimes, it takes days to get the data extracted. I have colleagues, friends in a project that I am working on, who basically have children with extended stays after birth with long records, and they cannot get the records to bring to the pediatrician, so the pediatrician can be aware. So, you hit right on it! How do we fix this? In other words, doctors will point to things like HIPAA and say, "Oh, I can't fax these records to you because of HIPAA laws." Total falsehood. It is not true. It is not just technology that we need. We need someone to say, you have a right to those records within 24 hours, and at this price, that should not be a state or federal issue; it should be an international issue. It is a human right, and I totally agree with you. How do you make that happen?

Dave deBronkart

You touched on a lot of really powerful things there. How do you make it happen? Get started for heaven's sake!

For another thing, I would recommend spreading the word about that Health Data Rights Declaration.

On the subject of that melanoma information going into that folder without being reviewed, you may have heard earlier this summer, there was a study, which announced that 1 in 14 lab results do not get communicated to the patient—7%! Now, imagine if, just take your head out of health care, and imagine if 7% of credit card transactions never ended up on a statement. If it is your credit card, you might be happy about this, but the merchant certainly would not. Here is the thing, at work, in my day job, where I still work part time, I manage the sales automation system called SalesForce.com. We can set up workflows that say, "after three days, send an email to somebody to remind them of this" or "when a new inquiry comes in, automatically send out a thank you email of the following form." These tools exist, and you know how much it costs to get started a big, hairy, automated system? Well, nothing. You go to salesforce.com, you click "free trial," and you start using it. That is cloud computing.

Not only that, the cloud can also securely back up your data. I was at a conference, the TEPR Conference in Palm Springs in February, and a high-level AMA (American Medical Association) official talked about the nightmare that happened when he installed his medical record system. Now, if somebody had tried to infiltrate a medical records conference with a horror story to defer adoption, this would have been a great way to do it. He lost all his data. The guy who manages his database could not get it back, and the following speaker said, "Don't you know about cloud computing? We have all our data stored in Google Docs and SalesForce.com. They do the backups for us." So, there are technologies out there.

Now, I will say, if you think back to when online banking first started, the early websites were ugly and hard to use, and there would be mistakes and issues. It took about 10 years to get it right and make it reliable. The clue is to get it started and certainly encourage each other to get off of the objection that it is too hard. It can be done.

Audience Member #2: "How to Change the Culture So Physicians Work Together With Patients?"

Hi, I am from Italy. This is the first time I am participating in Medicine 2.0. It is a wonderful venue, and actually, your talk was really engaging. I was very glad to hear your story. I would like you to comment on how participating in the decisions of medicine and physicians means a cultural change, which is possibly the most difficult thing to do. Actually, what happens most of the time, Dave, in my country, is, for example, the community of patients stay together and talk together about the potential therapies and alternative therapies, and they are very far from the decisions of their physicians. At the same time, physicians do not believe that they are looking for the right things, and the gap is growing. You said one very important thing is that the physician is a coach and I am a player. Actually, I think this is the key, but this is very difficult. How can physicians work together with patients to find the right solution? This is, I think, the key to get into the right process of patient participation in medical decisions.

Dave deBronkart: "Doctors Should Learn to Say 'I Do Not Know'"

That is a great question, and I want to make clear that I know there are idiots out there. There are idiot patients, and there are crooked doctors. Big surprise! I think that is true in every category of life. The way my physician, Danny Sands, expresses it is, "Embrace knowledge asymmetry, the idea that there can be knowledge on both sides of the equation."

In a nonmedical conference a year ago in San Francisco, somebody walked up to me, having heard that I had cancer, and said, "You know, chemo is a fraud. All you need is properly ionized water." I was like, "Well, thank you very much." So, although there are not known cases of death by googling, you can come up with some idiotic things.

Danny also mentioned at another conference that "in medical school, we are taught all kinds of things about how to have certain kinds of conversations with patients, but what we are not taught how to say is, 'I don't know. Let's find out.'" That is something that he is really good at and I know that. I know that personally, from my personal interactions with him. Is that a sufficient answer? I think also, by the way, the Society for Participatory Medicine is going to be working on developing advice and principles and practices for how to do that. So, you know, if you go sign up, you can participate in those discussions.

Audience Member #3: "How to Address the Lethal Lag Time"

My particular interest is in medical research and your statement about the lethal lag time: What do you think the solution is? Should we dump the traditional method of randomized controlled clinical trials in three years till publishing the results? Should we get rid of peer review or thinking radical ideas like that or just other methods to solve this electronically and reduce that lethal lag time?

Dave deBronkart: "What I Want Is Access to Useful Information When You Are Desperate"

That is a really good question because I know there are limitations to the peer-review process. I wrote a post last year about evidence-based medicine, which I had just learned about. Please understand the world that I live in. The good news is, for me, medically, I am now at the point where I only have a 50% chance or probability or the cancer coming back. There is a 50% chance or probability that I will be back there someday. I am doing everything I can in the way of attitude and finding information. But the fact is, my oncologist says, "We have no more advice for you on what you can do." So, here's a question. When you're at the fringes of medical knowledge, and the doctors have no answers, and your life is at stake, what do you do? Well, I'm not going to pitch out the whole peer-review process, where you have people who can say, "Wait a minute. This is a crock. This is poorly thought out and so on." At the same time, I know there are serious limitations to that. The reporting of relative risk reduction rather than absolute numbers is something I have learned about. Now, the Lipitor, on which we spend US $25 billion a year, actually affects one patient in 200 that takes it, if you get down to the raw numbers.

So, that is just the discussion of the peer-review process that I think is a good thing and needs to be understood for what it is and is not. On the other hand, what about when your life is at stake and the information has not been published yet? What I would like to see—Judy, you did something like this, right? Judy Feder, an astounding e-Patient woman who went and found unpublished information for a test that led to a treatment that has produced really favorable results.

So, I think, what I would really love to see is a very open database of all the studies that have been attempted or are in process, so that people with their physicians can get at those and find out what is out there, because if you are out of peer-reviewed stuff, you might as well look at the stuff that is not vetted yet.

A really good example of [communication problems regarding lag issues] is the following: The Kidney Cancer Association did a patient day two years ago in Cambridge, which was really so old school, old guard. What they did is they look at what doctors talk to each other about and tried to present it in patient terms. When question time came, I stood up and, at first, thanked the physician, because he was the head of the program at Beth Israel that had saved my life. I wanted to make sure he knew I was not just going to be doctor bashing. Then, I said, "If I understand it correctly—and this is based on stuff I learned from my patient community—the numbers you just put up there for median survival time and everything, those are from the Kansas City Kidney Cancer Base, correct?" He said, "Yes." I said, "If I understand correctly, all the data for that was collected ending in 1996, correct?" He said yes. "And at that time none of today's treatments existed, correct?" He said, "That's right." I said, "So, those numbers have nothing to do with somebody's expectations if they just got diagnosed today, right?" He said, "Well, that's correct." I said, "Well, I wish you would tell people that, you know, because you just scared half to death a guy at my table who's newly diagnosed!"

There is this discipline of saying, "What we will talk about is the latest peer-reviewed randomly controlled trial..." which may not be the most useful information. So, what I want is access to useful information when you are desperate.

Audience Member #4 (Judy Feder)

To your point, do you simply throw out the peer-reviewed research? No. However, I firmly believe that there is a place where evidence-based and experiential data have to meet and that we are smart enough and flexible enough, our brains are flexible enough, to figure out a way to negotiate that ground. I had a real epiphany a few months ago when I was telling somebody my story that I was being treated based on data that really was not "proven," and they said, "Well, this raises all kinds of regulatory and you know, liability issues and all kinds of things. You know, how did this happen?" I said, "Well, I didn't walk into my doctor and say, 'Hey, I would like Herceptin.'" The doctor said, "Yippee! Let's give you Herceptin." My doctor said, "Really interesting information. Let's see what the evidence says." So, that is the process that needs to go on. I mean, it needs to get vetted from multiple standpoints, but I think we are caught in this black-and-white thing that either it is what the doctors say from on high or it is the patient's report in their free-form discussions, and we are really trying to get to a medium.

Dave deBronkart

Yeah, that is absolutely perfect.

Audience Member #5 (Lisa)

Following up on what Judy was saying, one of the issues that you did not raise was that of the poor health literacy skills that far too many patients have, and I do not know what yours were to start with, but you developed wonderful literacy skills. However, for all the good e-Patients, there are also the many people who are looking for miracle cures and are very easily swayed by the far-too-abundant information that is out there that leads them in very negative paths, perhaps, not as Gunther showed, to death, but certainly to avoid appropriate medical care because they are following treatments. Working with Alzheimer patients, a lot of people follow advice, which is actually detrimental to their health, and certainly that is true with a lot of other diseases as well. However, it is also true around much simpler issues like weight loss, where people are looking for miracle cures and easy answers. So, I think that it is important to look at health literacy issues, and it is also important to look at the path that gets somebody to being an adept e-Patient.

Dave deBronkart

You know, Lisa, that is a great point, and it really goes to one of the most foundational issues in all of this. Alan Green from drgreen.com, who is the President of The Society for Participatory Medicine, is also quite a history student, and he has this array of Thomas Jefferson quotes that talk about this. From the founding of this country, well, that country over there—the United States—something about if the problem is that the people do not know enough to govern responsibly, the solution is to educate them, and at some point, you have to give people, like "Okay, I'm taking my hand off the bicycle. You're

on your own now. You know, you've got to make your own choices." For better or worse, that is the nature of democracy as well, and it has effects like the very weird things that are being said in the health care reform arguments in the United States now and so on.

It is a real issue, and it is why there is going to be a real need and real value in it. I know that is the work you do to help people achieve health literacy. Yes?

Audience Member #6

Thank you. I want to thank you for a really inspiring talk. Just to this point of coaching, there is a website called QuestionsAretheAnswer.com that was just published on Medicine on the net that helps patients communicate and ask the right questions to their doctor. There is a much older one that I am sure everybody here knows about, called "Quack Watch," that helps patients figure out what are the quack solutions out there.

I also wanted to ask about Tom Ferguson. I did not know he was ill. I met him in 1995 at a Leaders in Telemedicine conference, and he was an inspiring man.

Dave deBronkart

Yes, well, he had multiple myeloma. He was an actual e-Patient himself, and amazingly through unconventional channels, he learned that thalidomide, which had caused all those birth defects in the 1960s, was really useful for his condition, and it prolonged his life for years. He died unexpectedly in 2006.

Audience Member #6

He did. Oh, I am sorry to hear that.

Dave deBronkart

In an interesting twist, just yesterday, I found a link to an article. Apparently, it was a patient who had discovered that thalidomide could be useful for this and fed the information to the drug company, as a result of which, this product that had almost no market value started selling again. They quadrupled the price over the next few years. They have made another $300 million. I do not remember what the exact number is. They acknowledged her in their annual report but have not agreed to give her anything. So, she is now suing them for a portion of the extra earnings.

Audience Member #7

My last point is that kids are better at searching for information online than adults. Back in 1992, I started the first pediatric network online for kids here. You had a number of times where kids taught parents or adults with cancer where to look for the information.

Dave deBronkart

Terrific! That is great! That is great! See, we talk about the physicians who just do not want to use email, etc. Well, as time goes by, I know how that is going to unfold. Those people are retiring or dying and the others are coming up. So, you know, the iPod generation will take over.

Audience Member #8 (David Wiljer)

Hi, David Wilder from Princess Margaret Hospital here in Toronto. Thank you very much for your talk, but also for some valuable tools. We are doing a lot of work in the field of trying to give patients access to their data. One of the things we often hear is "What are patients actually going to do with it?" So, the next time I get that question, I am just going to send them to your e-Patients.net website. So, thank you for that.

You hit on a very important point though, and I think we need to reflect on that: Giving data in this way—what you are advocating for—I think requires a fundamental shift around a lot of different things, but fundamentally around traditional definitions of privacy, security, confidentiality, and a very fundamental idea of what it means for the Hippocratic oath of "we shall not do harm." For when we try to do this, we often hear clinicians say, "what we're very concerned about what harm we might do from this perspective." So, can you maybe give us a few thoughts on how we change this paradigm and how we are going to go about doing that? Thank you.

Dave deBronkart

Deep question, and it is one where I am not an expert. The issue of privacy and the risk of people having their hands on their data, it kind of gets back to the Jefferson subject of when people are new to something, they may make mistakes, but there is a learning process, and you focus on getting to the goal. I know HIPAA is a vast subject, and I also know that HIPAA is a mess because there are things like this story: David Kibby, who is a well-known physician, came back from a conference this summer and reported that he had heard from an attorney that somehow that woman had somebody else's data end up in her record, and the HIPAA goons at that hospital then locked her out of her own record because it contained data she was not supposed to see. Now, that is stupid.

At the same time, there was just a post last week on e-Patients.net by Susannah Fox about an article "Broken Promises of Privacy" from Paul Ohm, saying that the idea of "anonymized" data is actually a fallacy. Basically, smart hackers can de-anonymize data; the post starts by saying, "If you hate HIPAA, this is your lucky day," because real new evidence came along. The other thing is, the part of HIPAA that is supposed to give us our data does not work out well, because it can take up to 2 months and there can be a ridiculous charge for it. What I have started doing is asking for copies of my own records myself and putting them in my own system that is outside HIPAA. Ironically, one of the breakdowns in HIPAA, from what I have heard, again, I am not an expert, is that the way the law was changed in 2002, you may have given permission or your hospital may have given permission to partners of theirs to look at your data and do things with it without you knowing it. Ironically, your data may be more secure outside of that system. Personally, I am taking matters into my own hand. I am going to take responsibility for my own medical record. I am not sure how it will all play out, but I think we need something different than the way it works today.

Audience Member #9 (Peter)

Hi, my name is Peter. I too would like to thank you for the terrific presentation today. I just have two very quick questions:

First, has the integration between Beth Israel Deaconness' Health, you know, patient portal and Google Health improved since it began with you?

Second, maybe if you could just speak to the value for you, personally, of Google Health as a place to store information beyond Beth Israel Deaconness' circle of care?

Dave deBronkart

So, the answer to the first one is yes and no. Once it became national news, they did stop transmitting billing codes. Interestingly, it turns out that a bunch of people complained about that because when their well-managed billing codes can be useful at least as a starting point. It turns out that Microsoft Healthvault takes a whole different approach to it. With Google Health, you punch the button and your data is there and that is that, for better or worse. With Healthvault, everything that you send in from Walgreens or your hospital or whatever goes into a holding bin, where you look at it and inspect it and then you say, "Okay, send it over." Sounds to me like that is a much more sensible workflow. We do the same thing when we are putting a mailing list into our marketing system: We look at the stuff before we put it in because otherwise we end up spraying garbage on the walls, which, inside a computer, is messy.

Regarding the second question, the value for me, personally, of Google Health as a place to store information. If you go back and look at my April 1, 2009, blog post on e-Patients.net, what I see is, long story short, I want the power of innovation to be unleashed. However, innovative tools depend on data that they can work on. I mean, imagine, iPods would not have taken off much without digitized songs, and it is absolutely the same thing.

We are out of time, and I would again like to thank Gunther Eysenbach for inviting me to do this, and, as he said, as soon as that story hit the paper, he immediately grabbed me, and I am honored. Thanks very much to all of you.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Medicine 2.0'09 Opening Talk by Gunther Eysenbach (part 1).
[MP4 File (MP4 Video), 24987 KB - jmir_v21i11e17045_app1.mp4]

Multimedia Appendix 2
"Gimme my damn data!" - e-Patient Dave's keynote at Medicine 2.0 2009.
[PPTX File , 7242 KB - jmir_v21i11e17045_app2.pptx]

Reference

1. Eysenbach G. Medicine 2.0: social networking, collaboration, participation, apomediation, and openness. J Med Internet Res 2008;10(3):e22 [FREE Full text] [doi: 10.2196/jmir.1030] [Medline: 18725354]

Abbreviations

AMA: American Medical Association
ACOR: Association of Cancer Online Resources
HIPAA: Health Insurance Portability and Accountability Act
HIT: health information technology
IT: information technology
JMIR: Journal of Medical Internet Research
MIB: Medical Information Bank

Edited by G Eysenbach; submitted 13.11.19; this is a non–peer-reviewed article;accepted 13.11.19; published 22.11.19.

Please cite as:
deBronkart D, Eysenbach G
Gimme My Damn Data (and Let Patients Help!): The #GimmeMyDamnData Manifesto
J Med Internet Res 2019;21(11):e17045
URL: http://www.jmir.org/2019/11/e17045/
doi:10.2196/17045
PMID:31755873

Discussion Paper

Will Artificial Intelligence Translate Big Data Into Improved Medical Care or Be a Source of Confusing Intrusion? A Discussion Between a (Cautious) Physician Informatician and an (Optimistic) Medical Informatics Researcher

Qing Zeng-Treitler[1], PhD; Stuart J Nelson[1], MD

George Washington University, Washington, DC, DC, United States

Corresponding Author:
Qing Zeng-Treitler, PhD
George Washington University
2600 Virginia Ave NW
Washington, DC, 20037
United States
Phone: 1 (202) 994 2987
Fax: 1 (202) 994 2987
Email: zengq@gwu.edu

Abstract

Artificial intelligence (AI), the computerized capability of doing tasks, which until recently was thought to be the exclusive domain of human intelligence, has demonstrated great strides in the past decade. The abilities to play games, provide piloting for an automobile, and respond to spoken language are remarkable successes. How are the challenges and opportunities of medicine different from these challenges and how can we best apply these data-driven techniques to patient care and outcomes? A New England Journal of Medicine paper published in 1980 suggested that more well-defined “specialized” tasks of medical care were more amenable to computer assistance, while the breadth of approach required for defining a problem and narrowing down the problem space was less so, and perhaps, unachievable. On the other hand, one can argue that the modern version of AI, which uses data-driven approaches, will be the most useful in tackling tasks such as outcome prediction that are often difficult for clinicians and patients. The ability today to collect large volumes of data about a single individual (eg, through a wearable device) and the accumulation of large datasets about multiple persons receiving medical care has the potential to apply to the care of individuals. As these techniques of analysis, enumeration, aggregation, and presentation are brought to bear in medicine, the question arises as to their utility and applicability in that domain. Early efforts in decision support were found to be helpful; as the systems proliferated, later experiences have shown difficulties such as alert fatigue and physician burnout becoming more prevalent. Will something similar arise from data-driven predictions? Will empowering patients by equipping them with information gained from data analysis help? Patients, providers, technology, and policymakers each have a role to play in the development and utilization of AI in medicine. Some of the challenges, opportunities, and tradeoffs implicit here are presented as a dialog between a clinician (SJN) and an informatician (QZT).

(J Med Internet Res 2019;21(11):e16272) doi:10.2196/16272

KEYWORDS

artificial intelligence; big data; data driven approach; medical informatics; digital health; digital medicine; quality of care

Drs Nelson and Zeng-Treitler work together at the Biomedical Informatics Center at George Washington University. In the following we present a hypothetical dialogue which grew out of discussions they had as they considered their differing viewpoints of how artificial intelligence (AI) has developed and where it is going. While Dr Zeng-Treitler’s view of the future of AI is highly optimistic, Dr Nelson's opinion is more cautious. Dr Nelson was a practicing academic internist who became involved in informatics many years ago. He collaborated with Scott Blois on RECONSIDER (an early clinical decision support system) and on the Unified Medical Language System (UMLS) project. He eventually moved to the National Library of Medicine as Head of Medical Subject Headings. While at the National Library of Medicine, he fathered RxNorm, while continuing his work on UMLS and projects involving UMLS. Dr Zeng-Treitler has a background in computer science and

obtained her PhD in medical informatics from Columbia University. She has led a number of projects in clinical data mining, natural language processing, and consumer health informatics. During the past few years, her team has been actively investigating the use of AI techniques in clinical research including development of a novel explainable deep learning approach.

Dr Zeng-Treitler (The "Optimist"):

After decades of promises and disappointment, thanks to the seemingly unbounded computing resources and novel, data-driven methods, AI technology has finally arrived. From Jeopardy and Siri to face identification and autonomous vehicles, data-driven approaches have made the leap from laboratory experiments to applications that are transforming our lives outside health care. In some cases, these approaches have come close to passing the Turing Test—a test of a machine's ability to exhibit supposed human-like intelligence; machines can now perform some complex tasks such as image recognition and authentic game play as well as or better than humans would. Some would argue that the requisite approach is decidedly nonhuman. However, whatever the means to achieving these innovations, such successes have not been followed by analogous successes in health care.

One dramatic example of this disparity of accomplishment is AlphaZero, a computer game engine that mastered chess, Shogi, and Go. Even before the arrival of the current generation of data-driven artifacts, chess engines have been shown to be able to play at a level superior to that of chess champions. Players of Go (a board game that is thought to be much more complex than chess), however, believed that computers were no match for high-level professionals in this game. This belief was shattered first by AlphaGo, which soundly defeated the reigning world champion of Go. Then came AlphaZero. The news is no longer that such approaches can beat the champions of Go, chess, or Shogi. Rather, the remarkable fact is that AlphaZero did not learn from human experiences and that it defeated the best prior chess engines such as Stockfish. AlphaZero triumphed by playing more games against itself than had ever been played by all human players. This is not an approach we could readily duplicate in health care.

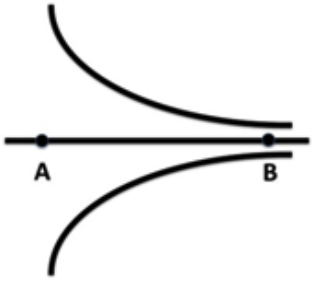
Dr Nelson (The "Cautious" One):

Is winning a game, with defined rules and objectives, really the best test of human intelligence? In hindsight, the answer is "No." For example, sophisticated chess playing programs have existed for nearly 50 years; from such programs, we learned how to organize computing resources to apply simple algorithms in a scalable way. Put differently, we did not learn anything about chess or how humans, even experts, play it. Instead, we learned that a supposed intelligence-requiring task was susceptible to a computational approach. We need to ask where and how such an approach is applicable in health care.

For example, when doing tasks that are generally thought to be human and creative, can the machine recognize when it is out of its depth? Sometimes, humans have the ability to do so. However, if we can define the realm closely enough, I agree that the machines can do wonders. So, how do we define the realm?

In Blois' seminal paper on Clinical Judgment and Computers [1], he described the world of a physician's thought process when seeing a patient, with the diagram shown in Figure 1. Point A for a physician would be where the patient walks in the door to be seen for the first time. The nature of the complaint, the context in which the complaint occurs, and all of the myriad possibilities are present. As the problem definition moves toward Point B, a computer is better able to manage the information and knowledge necessary for high-quality care. One way in which we can think about defining the realm is that we are moving toward Point B. Some computer scientists have argued that Point A is just about managing facts, but, as Blois observed, it is more about relevance—something that has proven difficult to replicate computationally.

Figure 1. The cognitive span required during diagnosis.



Dr Zeng-Treitler:

It is indeed important to define the realm for an AI application. Many tasks in health care are much more complex than game play, and we have not witnessed the triumphs of analogous approaches in the biomedical domain as have been achieved in game play. Quite a few studies have been applying the latest deep learning technology (a key AI method) to biomedical datasets [2-4]. The specific applications included image processing, natural language processing, and risk prediction. Deep learning, compared with traditional statistical and machine learning methods, has often shown modest improvements rather than breakthroughs [5-7].

Dr Nelson:

Whatever the details of these approaches, they apply nearly unbounded computing resources to very large amounts of data, something that has yet to happen in health care. Therefore, these approaches might prove helpful, but we do not know for sure yet.

For example, a simple question posed by a colleague is beyond our current capabilities: Given a patient who starts out with a feature of metabolic syndrome, which feature of the syndrome will he or she tend to exhibit next? Simplistically, this is exactly the kind of challenge that a data-driven approach should help with, and yet, it is currently "over the horizon" due to the insufficient data that were collected in the past.

Dr Zeng-Treitler:

Data are a key challenge when applying data-driven approaches to patient care. To begin with, biomedical data are highly complex. There many different types of data including image, text, numerical values, categorical classifications, and DNA sequences, representing tens of thousands of lab tests, procedures, diagnoses, medications, genetic markers, etc. Each data type also has its own characteristics; for example, a laboratory test value may need to be interpreted in the context of age, gender, and current conditions. However, diagnostic codes for different diseases have varying levels of accuracies.

In biomedical data analysis, there is also the paradox of having too much and not enough data at the same time. On one hand, there are a tremendous amount of medical record, social media, and literature data. Efforts like the Million Veterans Project [8] have also collected a huge amount of DNA data. Using devices for tasks like activity tracking and continuous glucose monitoring generates more data than our current medical record systems can digest. On the other hand, the health record of a patient is an open system with much missing information in contrast with the closed system of a chess or Go game, where all data are available. Patients are observed at irregular intervals (eg, at clinic visits or during hospitalization) and are never subjected to all possible tests or treatments. Sometimes, death is the only definitive outcome.

Dr Nelson:

I agree that data types are multiple and complex. Simple solutions are insufficient, and the proliferation of irrelevant data in a record, not to mention the current cut-and-paste or fill-in-the-template fad, obscures what is important.

One of the major difficulties with medical data is not just that it is not enough, but also that it is theory laden, that is, very few pieces of data are recorded routinely. A lot of data in observations are gathered only when the clinician has thought it is appropriate, that is, when diseases are tested for their absence or presence. If there is no reason to do the test, the test is not performed. Only a few tests are ever performed routinely; a transcribed set of physical observations (as is done in physical examination) is rarely recorded in enough detail (not to mention the failure to observe, which often occurs) to provide sufficient data for a more comprehensive analysis. For that reason alone, studies based on the recorded observations are often incomplete and potentially misleading. However, for predictions, observations not made may be the critical ones. Think about the patient with metabolic syndrome mentioned above. What data are we missing?

Similarly, the results of clinical trials are not a complete picture. Even though participants have been selected, often excluding many individuals because of complicating conditions, the data collected on the participants are designed for testing certain hypotheses, with narrowly defined outcomes. A common criticism is that such trials are so artificial that they are irrelevant.

Dr Zeng-Treitler:

The lack of integrated and standardized datasets is another issue. Although we can find many large datasets, they are often incomplete and difficult to link to other information. For example, environmental exposure, diet, physical activity, and genetic profile are among the common missing pieces of information when we examine the records about an individual patient. Detailed clinical trial datasets tend to lack long-term follow-up. Privacy issues and monetary incentives are also obstacles in data integration efforts.

Dr Nelson:

Real semantic interoperability to integrate and standardize datasets requires support in both terminology and how that terminology is used. Currently, human intervention is often required to interpret what one system is saying for use in another system. This situation is unfortunate; we can hope that over time, the necessary connections will take place (think of how the United States went from operator assistance on every telephone call to the automatic switching that takes place today). Such a change can only happen when many people see the need for and implement a common standard.

In addition, the notion of "large" in the context of health care data is only relative. Think, instead, of many experiments undertaken by Google; if they desire, the amount of data that can be used to develop and test a model is orders of magnitude greater than that available in health care.

Dr Zeng-Treitler:

In the domains where data-driven approaches have demonstrated success, there are outcomes that can be judged by human experts or machines themselves. For example, bilingual speakers can tell if natural language translation is working well, and the outcome of board or computer games can be easily determined. This allows easier simulation or annotation of data for machine learning. Such a task is much harder in the biomedical domains; investigation of causes or treatments of diseases in humans involve costly and long-term studies. In some cases, ethical concerns prohibit the experiments; for example, the introduction of potentially harmful genetic mutations into healthy human subjects is out of the question. We lack long-term outcome data for many treatments.

Dr Nelson:

I am not sure that there ever will be such a gold standard without a completely arbitrary definition. Variation between individuals is also a major obstacle. Although we perform studies using multiple subjects to account for biologic variability, our results are only approximate in their relevance to a given individual. For example, assuring genetic diversity in clinical trials is challenging, to say the least. Even the simplest tasks can be staggeringly multifactorial; for example, the information content of genetic testing for warfarin metabolism can be outweighed by whether the patient had lettuce for lunch.

To expand on this observation, suppose you have an automobile that is not working appropriately. Today, you consult the sensors and the computer readout to give you very precise information about what is going wrong. The automobile has a specific design, with specific parameters that can be measured. All of the vehicles of the same year make and model can be assumed to be alike in those important aspects. It is important to realize that every human (with the exception of identical twins) is genetically unique. In that way, people are very different from automobiles or other mechanical devices. To compound the complexity with which the cause of a human problem can be addressed, what the individual experiences throughout their life is unique. Although we have nice abstractions or methods of identifying individuals who share some common characteristics (whether the presence or absence of a disease, the response or lack thereof to a medication, the similar environment, or other considerations), these are only a shorthand notation. With 7 billion persons currently living in this world, the problem appears almost open-ended. Too often in data analysis, we look at diagnostic codes as having a deep meaning. These are accepted without any recognition of the degree of uncertainty of the diagnosis. All our data may be helpful and useful, but we need to continue to view them with a large grain of salt. The fact that Google Translate works as well as it does gives us hope, but as complex as natural language translation is, it is simpler than some clinical tasks.

Dr Zeng-Treitler:

Despite these challenges, applying data-driven approaches has the potential to transform health care. Today's health care is labor intensive, from scheduling and triage to diagnosis and treatment. Many tasks currently undertaken by humans can be carried out by intelligent software solutions supported by sufficient data. For instance, improved voice recognition and summarization technology might help reduce the amount of time patients and clinicians spend on paperwork. Improved decision support tools ought to be able to help patients decide about the appropriateness of seeking care. An accurate assessment of short- and long-term risks and benefits will inform treatment selection and lifestyle changes.

Dr Nelson:

To provide another use case, there is evidence that type II diabetes may be reversible, but it is hard to apply this knowledge to an individual patient. Given the patient in front of me, what should I do, or recommend, and with what expectation? Demographics, genomics, comorbidities, psyche, competing risks, and other medications, all play a role. How, in a given person, do I reconcile all the possibilities?

Dr Zeng-Treitler:

To develop these useful AI tools, we need better data, technology, and policy. To accumulate comprehensive, life-time data, patients must be in control and should be incentivized to share their data for research and care. Insurance, pharmaceutical, and medical institutions change over time. Currently, there are barriers for individuals to be the center of collecting the data on themselves. The barriers are present in data entry, collection, and storage; for example, some personalized health record products are tethered to an institution, while others require extensive transcribing efforts by patients or caregivers. Nevertheless, without patient consent and collaboration, gathering and linking longitudinal environmental, genetic, clinical, and behavioral data are neither feasible nor ethical. The current conditions are a huge barrier to any attempt to use data-driven approaches that have worked outside health care.

Efforts including PatienstLikeMe [9] and the All of Us Research Project of the National Institutes of Health [10] are examples of innovative approaches to curate bigger and better datasets. Most patients, however, are not engaged in such efforts. Patients are inherently motivated to improve their own health, but naturally have concerns regarding privacy and often do not see immediate benefits of participating in long-term studies. Appropriate incentives (eg, discounts for routine preventive care) coupled with security and authentication technologies are needed to entice a large and diverse population of patients to gather and share their data. The health care industry today owns parts of patient data and has limited motivation to purchase data from their customers. As the value of data increases, patients will become more valued as a partner.

Dr Nelson:

I agree that patients will need to assume the responsibility of carrying and sharing the information about themselves. However, experience tells us that not everyone is able or willing to do so. It will need cultural and political climate changes to encourage that development.

When we can collect data that are not directly what the philosophers would call "theory-laden," we may be able to refine our crude methods of patient diagnosis and care. I look forward to that day. If patients are the carrier of those data, it will be easier to obtain and use for analysis.

Dr Zeng-Treitler:

We also need to design and implement methods specifically for handling very large and "messy" clinical data. For example, we need to understand the context of missing data and errors to get a better picture of ground truth. A lab result may be missed because there is no indication for it, practice preference, an alternative method of assessing, or a failure of data entry. Imagine how much more difficult a chess game will be, if a human player or a chess engine could only observe some squares on the board at irregular time intervals with some error or distortion of the observation.

Further, we do not have an operational definition of "ground truth" in health care; a simple proposal is that one feature of ground truth is that it has predictive value—something that will be valued by clinicians and patients alike.

Dr Nelson:

Google has demonstrated that they can use lots of data to predict likely values for missing data in other areas [11]; however, it is yet to be determined whether this might work in medicine, but it is probably worth a trial. Irrespective of whether we can use large volumes of data to impute missing values, exploring how to handle the problem of absent observations is crucial,

especially whenever we try to apply the results of data-driven approaches to individual patients.

Another thought is that data that are missing, for any reason, are an observation in itself; the fact that the data were not obtained and recorded may be important. Think of the finding that the day and time of a test were more predictive of outcome than the result of the test [12]. We know that the data that are missing will have some predictive value.

Dr Zeng-Treitler:

On a different note, explanation of the data-driven models is critical to not only their adoption but also their impact [13]. Predicting that a patient will have certain adverse events in the next several days or years is desirable. It may be argued that it is even more important to know the modifiable factors that can reduce risk and enhance outcome. Since deep learning models can be highly nonlinear, we have the opportunity to discover novel and complex patterns.

Dr Nelson:

I agree that explaining the prediction is critical; it is something that separates health care from, say, recognizing whether an image is a dog or a cat. However, I think you meant to say predicting a patient will *probably* have some adverse outcome. Nothing in life is certain except that it will end. However, we can say "it appears this behavior or finding will likely have an effect on your future" and hopefully be able to express some degree of confidence in that prediction.

Learning how to express the confidence in a prediction is also important. How many folks really understand the statistics behind the predictions that occur today? What are the underlying assumptions behind any probabilistic model? It is more likely that with more frequent use and familiarity with the use of measures derived for AI models will lead to their acceptance.

Dr Zeng-Treitler:

I agree. These are all steps to be taken in order to optimize the use of big data through AI to improve medical care.

Dr Nelson:

As a parting thought, we need to be cautious about how intrusive data-driven approaches might be in the care process. Although McDonald [14] demonstrated that performance in care improves with reminders [14], the later experience has been one of too many reminders, leading to alert fatigue. When caregivers choose to override and ignore helpful information because of overload, have we accomplished anything?

I hope that careful design of systems and consideration of clinical workflow will alleviate the problem of excessive intrusiveness. Although it is tempting to just "let AI do it," the recent experiences with the Boeing 737 MAX demonstrate that there is danger in doing so. Neither AI nor a pilot alone is the optimal strategy in flying. In health care, involving patients more extensively in their care, together with AI and providers, may ultimately be an approach that works.

Conflicts of Interest

None declared.

References

1. Blois MS. Clinical Judgment and Computers. N Engl J Med 1980 Jul 24;303(4):192-197. [doi: 10.1056/nejm198007243030405]
2. Miotto R, Wang F, Wang S, Jiang X, Dudley JT. Deep learning for healthcare: review, opportunities and challenges. Brief Bioinform 2018 Nov 27;19(6):1236-1246 [FREE Full text] [doi: 10.1093/bib/bbx044] [Medline: 28481991]
3. Weng SF, Reps J, Kai J, Garibaldi JM, Qureshi N. Can machine-learning improve cardiovascular risk prediction using routine clinical data? PLoS ONE 2017 Apr 4;12(4):e0174944. [doi: 10.1371/journal.pone.0174944]
4. Miotto R, Li L, Kidd BA, Dudley JT. Deep Patient: An Unsupervised Representation to Predict the Future of Patients from the Electronic Health Records. Sci Rep 2016 May 17;6(1):26094 [FREE Full text] [doi: 10.1038/srep26094] [Medline: 27185194]
5. Quach K. The register. 2019. IBM Watson Health cuts back drug discovery 'artificial intelligence' after lackluster sales URL: https://www.theregister.co.uk/2019/04/18/ibm_watson_health [accessed 2019-10-23]
6. Rajkomar A, Oren E, Chen K, Dai AM, Hajaj N, Hardt M, et al. Scalable and accurate deep learning with electronic health records. NPJ Digit Med 2018 May 8;1(1):18 [FREE Full text] [doi: 10.1038/s41746-018-0029-1] [Medline: 31304302]
7. Lee J, Jun S, Cho Y, Lee H, Kim GB, Seo JB, et al. Deep Learning in Medical Imaging: General Overview. Korean J Radiol 2017;18(4):570. [doi: 10.3348/kjr.2017.18.4.570]
8. U.S Department of Veterans Affairs. Million Veteran Program (MVP) URL: https://www.research.va.gov/mvp/ [accessed 2019-10-23]
9. Patientslikeme. URL: https://www.patientslikeme.com [accessed 2019-10-23]
10. All of Us Research Program. URL: https://allofus.nih.gov [accessed 2019-10-23]
11. Halevy A, Norvig P, Pereira F. The unreasonable effectiveness of data. IEEE Intell Syst 2009 Mar;24(2):8-12. [doi: 10.1109/mis.2009.36]
12. Agniel D, Kohane IS, Weber GM. Biases in electronic health record data due to processes within the healthcare system: retrospective observational study. BMJ 2018 Apr 30:k1479. [doi: 10.1136/bmj.k1479]

13. Holzinger A, Langs G, Denk H, Zatloukal K, Müller H. Causability and explainabilty of artificial intelligence in medicine. WIREs Data Mining Knowl Discov 2019 Apr 02:e1312. [doi: 10.1002/widm.1312]
14. McDonald CJ. Protocol-Based Computer Reminders, the Quality of Care and the Non-Perfectibility of Man. N Engl J Med 1976 Dec 09;295(24):1351-1355. [doi: 10.1056/nejm197612092952405]

Abbreviations

AI: artificial intelligence
UMLS: Unified Medical Language System

Edited by G Eysenbach; submitted 15.09.19; peer-reviewed by A Holzinger; comments to author 14.10.19; revised version received 15.10.19; accepted 20.10.19; published 27.11.19.

Please cite as:
Zeng-Treitler Q, Nelson SJ
Will Artificial Intelligence Translate Big Data Into Improved Medical Care or Be a Source of Confusing Intrusion? A Discussion Between a (Cautious) Physician Informatician and an (Optimistic) Medical Informatics Researcher
J Med Internet Res 2019;21(11):e16272
URL: https://www.jmir.org/2019/11/e16272
doi:10.2196/16272
PMID:31774409

Original Paper

Leveraging Interdisciplinary Education Toward Securing the Future of Connected Health Research in Europe: Qualitative Study

Ioanna Chouvarda[1], PhD; Nicola Mountford[2], PhD; Vladimir Trajkovik[3], PhD; Tatjana Loncar-Turukalo[4], PhD; Tara Cusack[5], PhD

[1]Lab of Computing, Medical Informatics & Biomedical Imaging Technologies, School of Medicine, Aristotle University of Thessaloniki, Thessaloniki, Greece

[2]School of Business, Maynooth University, Maynooth, County Kildare, Ireland

[3]Faculty of Computer Science and Engineering, Saints Cyril and Methodius University, Skopje, the Former Yugoslav Republic of Macedonia

[4]Faculty of Technical Sciences, University of Novi Sad, Novi Sad, Serbia

[5]Health Sciences Centre, School of Public Health, Physiotherapy and Sports Science, University College Dublin, Dublin, Ireland

Corresponding Author:
Tara Cusack, PhD
Health Sciences Centre
School of Public Health, Physiotherapy and Sports Science
University College Dublin
Belfield, Dublin 4
Ireland
Phone: 353 1 7166515
Email: t.cusack@ucd.ie

Abstract

Background: Connected health (CH) technologies have resulted in a paradigm shift, moving health care steadily toward a more patient-centered delivery approach. CH requires a broad range of disciplinary expertise from across the spectrum to work in a cohesive and productive way. Building this interdisciplinary relationship at an earlier stage of career development may nurture and accelerate the CH developments and innovations required for future health care.

Objective: This study aimed to explore the perceptions of interdisciplinary CH researchers regarding the design and delivery of an interdisciplinary education (IDE) module for disciplines currently engaged in CH research (engineers, computer scientists, health care practitioners, and policy makers). This study also investigated whether this module should be delivered as a taught component of an undergraduate, master's, or doctoral program to facilitate the development of interdisciplinary learning.

Methods: A qualitative, cross-institutional, multistage research approach was adopted, which involved a background study of fundamental concepts, individual interviews with CH researchers in Greece (n=9), and two structured group feedback sessions with CH researchers in Ireland (n=10/16). Thematic analysis was used to identify the themes emerging from the interviews and structured group feedback sessions.

Results: A total of two sets of findings emerged from the data. In the first instance, challenges to interdisciplinary work were identified, including communication challenges, divergent awareness of state-of-the-art CH technologies across disciplines, and cultural resistance to interdisciplinarity. The second set of findings were related to the design for interdisciplinarity. In this regard, the need to link research and education with real-world practice emerged as a key design concern. Positioning within the program context was also considered to be important with a need to balance early intervention to embed integration with later repeat interventions that maximize opportunities to share skills and experiences.

Conclusions: The authors raise and address challenges to interdisciplinary program design for CH based on an abductive approach combining interdisciplinary and interprofessional education literature and the collection of qualitative data. This recipe approach for interdisciplinary design offers guidelines for policy makers, educators, and innovators in the CH space. Gaining insight from CH researchers regarding the development of an IDE module has offered the designers a novel insight regarding the curriculum, timing, delivery, and potential challenges that may be encountered.

(J Med Internet Res 2019;21(11):e14020) doi:10.2196/14020

KEYWORDS

connected health; interdisciplinary studies; interdisciplinary research; problem-based learning; technology and health

Introduction

Background

The population of Europe is aging. Life expectancy is estimated to increase by more than a full year between 2016 and 2021—from 73 to 74.1 years—bringing the number of people aged over 65 years to more than 656 million or 11.5% of the total population [1]. Coupled with the increased incidence of chronic disease, this illustrates the challenges being faced by the health care system. Currently, hospital admissions consume more than 37 million bed days each year across the European Union (EU). However, the digital transformation of health and care, a key component of the EU's Digital Single Market, offers tremendous potential for improving the prevention, detection, and management of chronic diseases as well as improving health system management and research [2]. New and innovative ways to maintain and enhance health are required. Connected health (CH) describes the use of technology to provide health care services in a more flexible and cost-effective way for both citizens and health care practitioners. CH refers to [3]:

> *A conceptual model for health management where devices, services or interventions are designed around the person's needs, and health-related data is shared, in such a way that the person can receive care in the most proactive and efficient manner possible. All stakeholders in the process are "connected" by means of timely sharing and presentation of accurate and pertinent information regarding patient status through smarter use of data, devices, communication platforms and people.*

CH can be used by both clinicians and citizens to enable better and more efficient use of scarce health care resources. It promises a paradigm shift with novel technologies being used to create and develop links between individuals and communities, health and disease, and different health actors (eg, citizens, patients, clinicians, and policy makers), thereby enabling citizens and health care practitioners to make better decisions about health care.

CH interventions such as home-based exercise programs can contribute to CH impact via improvement in patient adherence [4], but they require reshaping of health care. Therefore, motivating stakeholder involvement to embrace such change requires some existing evidence on CH effectiveness and advantages. Unfortunately, there is still some distance to go in terms of proving the effectiveness of such programs [5]. The challenges in achieving such evidence-based proof are both technological and medical. The role of digital health education and the resulting literacy of CH-involved stakeholders have been highlighted [6], especially with regard to quality and safety concerns in health data as well as the fear of unintended consequences of technology use. A comprehensive set of evidence-based guidelines for CH might allow it to achieve its full potential [7]. Working toward the necessary evidence and medical guidelines will, however, require the development of a new CH culture that is actively embraced by CH stakeholders across multiple domains and disciplines.

Adopting a CH approach to health care could enhance the delivery of a more integrated health care system [8]. For health care to be *integrated*, it must connect "inputs, delivery, management and organization of services related to diagnosis, treatment, care, rehabilitation and health promotion" [9]. Such coordination of care requires, however, a corresponding coordination of multiple professional camps, such as clinicians, medical device engineers, health care managers, and policy makers. CH can offer a key building block in this regard, offering the tools to coordinate consistent care across settings and over time and bringing together the various groups involved in patient care to decide, organize, and deliver services [8].

Concepts central to both the connectedness and integration aspects of the CH approach are sharing of information, involvement and collaboration of multiple professions, redesign of care models, temporal and organizational continuum, new care and business models, and leveraging of services via technology. The development of CH technologies, therefore, requires that professionals from across both the health care and technology disciplines work together in a more integrated and cohesive way. This implies a good understanding of the skills each discipline brings to the team and the health needs they intend to address [9]. Furthermore, these new services and care models involve innovation, which, in turn, involves teamwork and synthesis, including also a closer integration between health and biomedical sciences, as well as information and communications technology (ICT) sciences [10]. A natural driver for such developments is education. Some work has already been undertaken in this area, and the European Network for the Joint Evaluation of Connected Health Technologies (ENJECT) report [11] refers to education programs across Europe and the extent to which they are creating the conditions and skills necessary for the widespread adoption of CH. It is essential to educate the next generation of researchers, clinicians, scientists, and decision makers about the importance of CH so that they, in turn, will be able to empower citizens to engage with CH as fully as possible through their own professional lives.

Medical students are envisioned as the frontline of *digital natives*, and telemedicine education is proposed for their preclinical and clinical curricula [12]. Recognizing that telemedicine and telehealth will play an increasingly essential role in the delivery of health care, the American Medical Association has called for telehealth to become a core competency of medical students. They do so in the knowledge that studies have demonstrated that a principal reason why physicians do not practice telehealth is lack of education despite patients' interest in telehealth [10].

Slovensky et al [13] proposed a set of mobile health (mHealth) skills for health professionals as a necessary enhancement for clinical training programs that ultimately would benefit both providers and patients. These included (1) digital communication

skills, (2) technology literacy and usage skills, (3) mHealth products and services, (4) regulatory and compliance issues, and (5) the technology business case. In the CH context, two key areas of education have been identified for future nursing graduates: (1) the actual use of new technologies and (2) managing and making sense of the data produced (data literacy to interpret and make use of ambient data) [14]. These are clear indicators of the need to consider CH education in different health care professions.

However, CH involves a broad spectrum of disciplines extending beyond those directly associated with traditional health care and health care delivery. Recent studies and research programs have begun to recognize that engineers, computer scientists, information technology designers, and policy makers must together engage in interdisciplinary research to deliver on the promise of CH. Xu et al [10] emphasize the need for the discovery of new diagnostic tools and treatments by using a multidisciplinary and highly collaborative approach. Mountford et al [15] extend this to the education of CH researchers, presenting the need for comprehensive training and research program by embracing all key elements—technical, social, and economic sciences—required to produce researchers and project outcomes that are capable of meeting existing and future needs in cancer rehabilitation. We argue that relevant professionals outside of the health care field (beyond junior researchers) also need to understand the nuances and requirements of CH within the domain of modern medicine. For all disciplines to participate in the development of CH and to ensure that innovations are fit for purpose, it is imperative that all disciplines work together to enable a more informed and potentially creative future. It is, therefore, apparent that CH interdisciplinary team working needs to be established both in day-to-day practice and in the education programs that prepare professionals for such practice.

The Research Question

The main premise underpinning this research is that modeling interdisciplinary team working behaviors during the education process is a prerequisite for productive interdisciplinary cooperation in CH research and development as well as practice. Clear understanding of benefits, potential, and relevance of interdisciplinary collaboration must be learned and practiced to yield its full potential in the field. Education of health care professionals should embrace the technological innovation that is empowering the field of medicine today. Engineering curricula, being far more flexible in continuous adaptations to the exponential progress in the field, have already encompassed enabling technology for health. However, only bringing the interdisciplinary teamwork into the classroom will ensure broader acceptance and understanding of the fundamental CH concepts. Investment at the education level will eventually facilitate faster adaptation and implementation of CH technologies. The ENJECT Cooperation in Science and Technology (COST) action [16] initiative to summarize the existing readiness of European countries to adopt CH included an investigation into CH education curricula and revealed disappointingly modest efforts toward CH throughout third-level education [10]. The survey presents responses across 15 European countries to questions that assessed the prevalence of university-based programs that educate and equip health care professionals or future health care professionals to engage with CH. The results showed a reliance on biomedical informatics courses, with less than half of the sample programs offered by respondents having a CH or electronic health dimension. More than half (22 out of 42) of the programs cited had health care informatics as a major component. The majority of these 22 programs included the phrase *health informatics* or *information management* in their titles. This snapshot of health professional education indicated a lack of progress in the incorporation of CH into their program content, leading us to question the broader interdisciplinary context around CH education. If health care professional educators were struggling to span the CH interdisciplinary divide, then how might other relevant disciplines be faring in this regard? Given the inherently interdisciplinary nature of CH, we set out to qualitatively develop deeper understandings of the challenges that might underpin the ENJECT findings. Although core elements of an interdisciplinary program at the fourth level (PhD) have been proposed based on the experience of Marie Curie Actions for funding early-stage researcher programs [17], such context-specific methods for interdisciplinary education (IDE) cannot simply be transferred to all levels and contexts of academic education without more research and adaptation. This paper addresses that deficit by exploring how third-level IDE in the domain of CH might be leveraged to secure future development and practice of CH. Important questions concerning CH education are raised in relation to the disciplines involved, professional boundaries, crossing boundaries, and the optimum ways in which to deliver interdisciplinary CH education.

With some notable exceptions, previously presented, existing literature offers few guidelines as to the design of IDE modules for CH teams, no advice on how much interprofessional education (IPE) or IDE is adequate, and no indication of how its objectives should be defined. The aim of this study was to examine the perceptions of researchers engaged in interdisciplinary research regarding the design and development of CH education modules.

Related Work

Interprofessional Versus Interdisciplinary Education: Fundamental Concepts

IPE and multiprofessional education are examined in the literature [18]. There is some discrepancy in the literature regarding the terminology used to describe bringing various disciplines or professions together to learn. IPE is defined as occasions when students of two or more professions learn with, from, and about each other to improve collaboration and the quality of care and services, according to the Centre for the Advancement of Interprofessional Education [19]. IPE is well established in health care education. The learning objectives, as expressed in the study by L'Ecuyer [20], are (1) to communicate the professional roles and responsibilities of all team members clearly to others, (2) to understand the relationship between effective team communication and improved patient safety and health outcomes, and (3) to demonstrate skills of effective interprofessional team and patient-centered communications that integrate the knowledge

and experience of other health professionals and patients to provide appropriate care. Although well established in the health professional sphere, IPE does not often encompass key nonclinical members of the contemporary health care team, such as medical engineers and informaticians or medical data analysts. Although the term has traditionally been defined solely in the context of health care professions, we propose that IPE now needs to be seen in a broader context. If CH research is to achieve its full potential, then there is a need for IPE to embrace professionals beyond health care—in particular engineers, computer scientists, physicists, and mathematicians who have a crucial role in designing the technologies required for CH. IDE, on the other hand, is defined as [21]:

> *An interaction involving collaborations between students from differing subject areas in pooling their disciplinary knowledge in addressing complex and significant, real-world problems [leading to] the ability to understand and be understood by a diverse group of specialists.*

Smith and Clouder [22] explored the similarities and differences between interprofessional and IDE. As mentioned, definitions of interprofessional learning tend to link more directly to practice and the workplace than definitions of interdisciplinary learning. However, both explanations stress the centrality of collaboration and integration toward addressing complex problems. Emphasizing the synthetic procedure in the study of Aboelela et al [23], interdisciplinary research integrates the analytical strengths of two or more often disparate scientific disciplines to solve a given biological problem. While engaging in this mission, the terminologies, approaches, and methodologies may be gradually merged, and the scope of investigation may broaden and may even lead to new hybrid disciplines.

Overall, within an interdisciplinary team, *each team member builds on each other's expertise to achieve common, shared goals, for example, toward integrated care.* In this paper, between IPE and IDE, the latter term is adopted.

Challenges, Barriers, and Descriptors of Interdisciplinary Work

Indicatively, IPE barriers have been identified in a study by Hall [24]: each health care profession has a different culture, which includes values, beliefs, attitudes, customs, and behaviors that contribute to the challenges of effective interprofessional teamwork. In this work, a clear and recognizable idea or goal, serving as the focus for team members, is suggested as an opportunity for addressing such barriers for teamwork to succeed. This *idea dominance* allows each member to shift from their specific professional focus to one requiring an understanding of another's observations and interpretations. Problem-based learning (PBL) for IDE and IPE can support this idea and can be a vector of success in the context of CH.

Beyond IPE among health professionals, CH training involving all relevant disciplines is not well explored. Attitudes toward interdisciplinary training have been studied where staff and students have been drawn from medical and engineering backgrounds. In a study by Spoelstra et al [25], medical and engineering students and staff attitudes were examined, and important differences were reported for staff and students between the disciplines regarding attitudes toward and perceptions of the relevance of interdisciplinary learning opportunities and the role of creativity and innovation. There was agreement across groups concerning preferred learning, instructional styles, and module content. Medical students showed greater resistance to the use of structured creativity tools and interdisciplinary teams. Such attitudes could be dealt with early in an educational program. As mentioned in the study by Feyerabend [26], viewing science too ideologically and rigidly, similar to a religion, and becoming dogmatic impair the overall progress of science.

Mountford et al [27] comment on the increase of interdisciplinary research networks at the doctoral research level to increase innovation, creativity, and knowledge and focus on three such CH doctoral research networks that have been funded by the EU. They raise concerns as to the structuring of these networks to accomplish both deep disciplinary goals and broader interdisciplinary objectives at the same time. On the basis of 28 semistructured interviews with the doctoral students on these programs, they outline three key elements to enhance the development of interdisciplinary social capital within such networks: structuring the program to facilitate the extraction of value for each student from the interdisciplinary process, motivating students throughout the interdisciplinary program journey, and facilitating students to relate to others both within and external to the program.

From another perspective, the value of interdisciplinarity in research has been critically examined [28]. Various researchers have explored the value of interdisciplinarity in terms of citation and funding. In a study by Bromham et al [29], it was found that the greater the degree of interdisciplinarity, the lower the probability of being funded, whereas in the study by Larivière et al [30], it was found that distance in interdisciplinarity increases scientific impact of publications.

These studies introduce descriptors of interdisciplinary research [31], having as a basis the work of Stirling [32,33]. In that work, *variety*, *balance*, *and disparity* were introduced as indicators of disciplinary diversity. *Variety* refers to the number of disciplinary categories, *balance* is related to the evenness of the distribution of disciplines, and *disparity* measures the extent to which these disciplines are different or similar from a cognitive point of view. Such descriptors may be useful in describing necessary or typical interdisciplinarity in CH education and research.

Paradigms of Interdisciplinary Education

IDE assumes innovative teaching methodologies, as it should foster active students' involvement, exchange of opinions, and cooperation. The most widely used paradigms for IDE include competency-based learning, PBL, project-based learning, and design-based learning. Competency-based learning builds students' knowledge, one competency at a time. In that sense, it is a natural way to introduce different disciplines from along the continuum of learning into the students' body of knowledge. The formative method of assessment used in competency-based learning places an emphasis on the application of knowledge

in a certain situation (problem), focusing this educational paradigm on the skills acquired by the students [34]. As specific individual skills might be more challenging to obtain for different students, the collaborative learning approaches can be used to provide more flexible learning environment. The work done by Hall Barber et al [35] presents the results of the introduction of competency-based learning approach in medical studies. To enable students from different disciplines to work together, *PBL* has been used in IPE for health care professional students [36]. It has been considered as a means of encouraging self-directed learning, critical thinking, lifelong learning, and self-evolution among students. *Project-based learning* [37] involves the solution of a problem set by the student or instructor. This question or problem in focus serves to organize and drive activities toward a *solution that addresses the driving question*. It involves initiative by the student or group of students, and a variety of educational activities constitute parts. It usually results in a product (eg, a report and a computer program) delivered after a considerable length of time and investment of work effort. Teaching staff only play a facilitatory role in the learning process.

When considering project-based learning versus PBL, the starting point in both approaches is a problem; however, in PBL, students' activity is directed to *studying*, whereas in project-based learning, students' activity is directed to constructing the solution or product [37].

Design-based learning has recently been proposed [38] as a means to help to bridge the gap between research and practice in medical education because it contributes toward both theory testing and refinement on the one hand and improvement of educational practice on the other hand. This genuinely introduces interdisciplinarity [39]. Its main aspects are (1) iterative cycles of design, evaluation, and redesign; (2) authentic real-life learning settings; (3) testing and refining theories as well as advancing practice; (4) mixed methods studies; and (5) interaction among designers, researchers, and practitioners with different expertise.

Project-based learning and design-based learning methods seem to present differences, as discussed in the study by Stokholm [40], not only in the procedure of learning but also in the foreseen competence creation. The former leans more toward discursive thinking and an analytical-oriented working mode, whereas the latter toward design and innovation theories, methods, and tools as well as a culture of systemic thinking and a synthetic-oriented working mode.

Methods

Overall Methodology

To address the aim of this work and explore academic IDE toward the evolution of CH, a cross-institutional (Aristotle University of Thessaloniki [AUTH], Greece, and University College Dublin [UCD], Ireland), multistep qualitative approach was defined. Although these institutions are geographically distant, they are similar in many respects. Both universities are state funded, and they both have highly competitive, long-standing, and sought-after professional health education programs (nursing and medicine) and computer science and engineering programs. They also have well-developed collaborative interdisciplinary postgraduate research education programs, which have a CH ethos at their core. The AUTH Lab of Computing and Medical Informatics and Biomedical Imaging Technologies (MI-LAB) and the UCD Insight Centre for Data Analytics are international leaders in CH research and have collaborated in several EU-funded CH projects, for example, the initial training network "Connected Health Early Stage Researcher Support System" and the project "Championing a Multi-Sectoral Education and Learning Experience to Open New Pathways for Doctoral Students."

The data collection was conducted in two stages:

- Stage 1: interviews with CH researchers to explore specific themes
- Stage 2: workshops to further examine and elaborate on themes emerging from stage 1

Stage 1 research was undertaken in AUTH, whereas stage 2 data collection was undertaken in UCD.

Stage 1

Interviews were undertaken with researchers in the health informatics sector collaborating with the AUTH MI-LAB, Greece. The purpose of the interviews was to gain insight regarding the views of researchers about the needs for CH education and the ways in which CH education could be best delivered. Ethical review was not required as this study was the evaluation of standard educational practices. All participants consented to participate and were assured of confidentiality and anonymity. The interviews were undertaken by an experienced qualitative researcher who has published several papers that have employed similar research methodologies [27,41-43].

All staff and postdoctoral and doctoral researchers employed in or closely collaborating with AUTH MI-LAB in health informatics research or education were invited to participate in the study. A total of 9 individuals agreed to be interviewed, comprising 2 staff and 7 early-stage researchers (at the postdoctoral or PhD level). The majority of those interviewed were engineers or computer scientists, who had specialized in health informatics. All those interviewed had undertaken their research in an interdisciplinary context and had experience of communicating with a broad range of disciplines. All researchers had contributed to the health informatics module delivered to the undergraduate medical students at Aristotle University. The interview schedule is presented in Multimedia Appendix 1. All interviews were audio recorded, transcribed verbatim, and examined for emergent themes according to the method described by Braun and Clarke [44]. The technique of thematic analysis, as described in the study by Braun and Clarke, was used to analyze the interview data. Thematic analysis enables the identification, analysis, and reporting of patterns or themes that occur within a qualitative dataset while also offering a robust method of organizing, describing, exploring, and analyzing the data [45]. The interview transcripts were initially examined through the process of reading and rereading (TC and NM). The transcripts were read independently to enable the researchers to become familiar with the data. Each person then

made individual observations in relation to the interview content while developing preliminary inductive codes. Working together, the authors developed a coding framework, which was subsequently used to examine the data. To ensure consistency, the authors (NM and TC) discussed agreements and disagreements to reach a consensus regarding the emergent themes. The themes were aligned with the codes, thereby enabling the development of a narrative. Through the analysis process, the perceptions regarding how interdisciplinary could be designed and delivered were examined. Qualitative data analysis software was not used in this study.

Stage 2

This stage included workshops to gain further knowledge and insight. These workshops took place at the Insight Centre for Data Analytics, UCD, Ireland. The method used in both workshops was based on the structured group feedback approach [46]. This method was selected to garner participant responses in relation to the themes that emerged in stage 1. In all cases, the participants were informed and consented to the use of their data for research purposes, and no personal data were collected.

Workshop 1 was entitled "Finding our way for interprofessional connected health education" and followed a 3-phase procedure. In phase 1, participants were split into *education* and *teamwork* groups. Those participating in the education group had to individually reflect on four questions and answer in written form: regarding interdisciplinary educational examples; the education level for a CH course; benefits, challenges, and objectives for a CH course in an interprofessional class; and potential project setups for a CH class. Similarly, the CH teamwork group had to reflect on and write answers on interdisciplinary teamwork examples, potential concept and terminology or language barriers in interdisciplinary work, interdisciplinary teamwork barriers in CH research, and potential beliefs and attitudes to be addressed in CH teamwork. The questions are listed in Multimedia Appendix 2. In phase 2, the group members came together to discuss their answers and reach a fusion of ideas as well as a consensus or a ranking and prioritization within group. In phase 3, a plenary discussion took place. Each group rapporteur presented their results to the whole group of participants, and a final consensus was reached.

This workshop had 10 participants, all of them researchers in CH, including postdoctoral researchers (n=3), PhD students (n=4), and professors (n=3), from medicine, physiotherapy, business, social sciences, and ICT backgrounds. The duration of the workshop was 2 hours.

Workshop 2 took place during a full group meeting of CH researchers at the Insight Centre and included 16 participants, including postdoctoral researchers (n=4), PhD students (n=10), and professors (n=2). The disciplines represented were the same as those of the first workshop. The duration of this second workshop was less than 2 hours. Building on the outcomes of workshop 1, the purpose of this workshop was to refine and elaborate on specific points. Therefore, the previous procedure was followed, but with the addition of two new questions that focused on preconceptions in an interdisciplinary team that could be addressed through education and the focus of an interdisciplinary CH course (see Multimedia Appendix 3), and again, the participants were split in two groups that answered and discussed the two questions before coming back to a full group discussion.

Results

General Findings

The analysis of the interviews revealed a number of key themes as follows: challenges concerning IDE, recommendations in relation to developing IDE, and positioning of IDE within a curriculum (Table 1). Quotes were chosen that were illustrative of the subthemes and themes emerging, with due regard to representation across institutions, disciplines, and career stages. The interviews were complemented by the workshop results on CH education and interdisciplinary teamworking.

Table 1. Emergent themes and subthemes from the interviews undertaken at the Lab of Computing, Medical Informatics and Biomedical Imaging Technologies, Aristotle University.

Theme	Subthemes
Challenges identified to interdisciplinary learning	Communication, state-of-the-art knowledge, and resistance to interdisciplinarity
Recommendations for an interdisciplinary learning environment	Learning environment and link to real-world practice
Position of the interdisciplinary module within the study program	Need for skills and experience and early integration

Interview Results From the Lab of Computing and Medical Informatics and Biomedical Imaging Technologies, Aristotle University

Challenges Identified to Interdisciplinary Learning

The researchers interviewed in health informatics domain at the Aristotle University identified several key themes in relation to CH and CH education. They identified several challenges that centered on working across and between disciplines. The challenges were frequently associated with communication and the difficulty experienced in communicating their own key disciplinary concepts while also understanding those of an individual from another discipline in an interdisciplinary context. They identified that each discipline perceives ideas and concepts differently and that it can take time to convey these concepts to an individual beyond their own discipline. One researcher commented as follows:

> *Difficult because we cannot communicate properly our needs both of us...I ask something my own words because I have something on my mind and I say because it is a different way of thinking so we have*

> *to sit down and discuss for a long time so that we can understand each other.* [AUTH 2]

The fast pace of technological advancement is beyond doubt; however, some concern was expressed regarding the up-to-date understanding of members of other professions within the team. One participant commented:

> *The problem is with the older doctors who think technology is not for them and try to put the burden on you...they too must understand things so that they can help me and their selves.* [AUTH 6]

The interview participants frequently referred to their links with health care staff, their need to work together, and the importance of being able to communicate and collaborate as being an essential element of their work:

> *...constantly in communication with the health care professionals in order to explain to them how to explain to the patients how to use our devices and our applications.* [AUTH 5]

A number of participants raised the possibility of resistance from other disciplines to engage in an interdisciplinary context:

> *I don't know the level of willing that's let's just say from different departments to join forces.* [AUTH 1]

Recommendations for an Interdisciplinary Learning Environment

The staff who were interviewed had considerable experience in terms of education delivery, many of whom had delivered lectures and supported the laboratory lessons for the undergraduate medical program and curriculum. When they were asked to describe how best they thought IDE could be delivered, several critical recommendations for learning emerged.

A number of participants recommended that students should learn to work as part of a multidisciplinary team from early in their education, as they believed this would prepare them to collaborate and communicate with disciplines beyond their own from early in their careers:

> *...laboratory sessions for students...team orientated...multidisciplinary projects for the students in order to learn from early on how they can collaborate and how they can speak the same language.* [AUTH 1]

A number of participants identified the importance of learning in an enjoyable environment because they believed that this promotes better engagement and an opportunity for students to extend their learning beyond their own disciplinary limits:

> *...have a little bit of fun a play…play that will turn on your imagination not for hard work…that you are doing daily…let's have fun and be out of the box you know...* [AUTH 3]

One person identified the value of interdisciplinary work as a means of empowering students to link research and clinical outcome:

> *More practical things how to link the research with the clinical outcome.* [AUTH 4]

Another indicated that interdisciplinary work is an opportunity to innovate in terms of education and enable the introduction of contemporary topics with a view to stimulating student interest in engagement and learning:

> *...cutting edge and innovations which are currently hot in the science field.* [AUTH 5]

Position of the Interdisciplinary Module Within the Study Program

The interview participants made some recommendations in terms of the position of an interdisciplinary module within a program. A number of participants commented that owing to the wide variety of career pathways that can be chosen by electrical engineers and computer scientists, many of which do not involve engagements with health care, it was thought that offering such an opportunity late in an undergraduate program or as part of a master's degree program would be most appropriate One participant commented as follows:

> *...only at the last years when you have chosen your faculty...because electrical engineering is versatile...its only one tenth of what engineering can do.* [AUTH 6]

Another participant indicated that introducing such innovations at an early stage within the formalized education cycle might be difficult; however, it could be developed initially as an extracurricular activity:

> *...in order to include this seminars this lesson in the everyday curriculum of high school or the first semester of college or university I think this is a little bit difficult all this could be in extracurricular.* [AUTH 5]

Workshop Results From the Insight Centre for Data Analytics, University College Dublin

With respect to existing or potential cases of IDE, quite a few education examples emerged—"Patient care in the long term" or a "CH design project" (similar to a weight management program design). These cases were distinguished from other more clearly multiprofessional education examples such as biotechnology programs, patient care in *case* discussions within medical training, and training in rehabilitation programs. In these multiprofessional education examples, the role of multiple health professionals appears, but rather as forming parallel and not fusing or interacting paths. Overall, it was more challenging to find interprofessional courses.

With respect to positioning of the module within the program, two levels were identified for CH education with two different aims: (1) at undergraduate level for awareness-raising purposes, not necessarily focusing on interdisciplinarity, and (2) at postgraduate level for actual interdisciplinary research. An alternative idea was to embed elements within different courses at the undergraduate level. At the undergraduate level, it was found essential to inform students about the roles of different actors and to address misconceptions, perhaps by presenting examples of how interprofessional projects worked. To build a

common language within an interdisciplinary PBL team, participants believed that some time has to be invested, and each member has to have at least an undergraduate level of proficiency at their discipline. This indicates the master's level as an appropriate time for achieving the second aim of CH education. At later stages, it was felt that the lack of time, professional duties, and professional bias might hinder the potential involvement in interdisciplinary research teams.

Project-based and problem-based courses were discussed as a means to involve young researchers, with *fresh eyes* in interdisciplinary experience. Course setup examples were mentioned. For project-based learning, such course examples included health and well-being solutions or CH projects, whereas PBL examples included patient care at large as a topic and a hackathon as the instrument. Learning should take place based on real-life scenarios and solutions or actual clinical studies and authentic problems that have not been accommodated under traditional care scenarios. Participants suggested learning in a practical way, using real products where clinicians can see outcomes and by locating real problem-based situations and investigating how technology can solve these problems. Participants were adamant that the PBL experience should include learning to map the problem, as designers do.

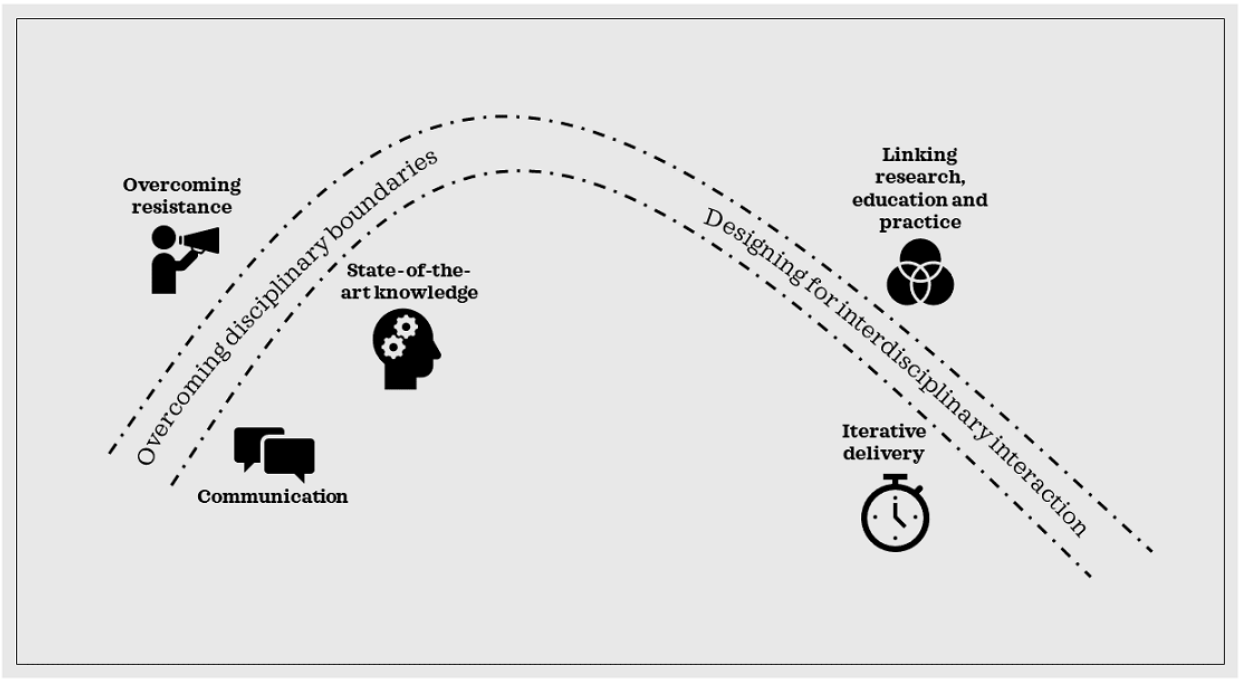
In general, teamwork learning was placed as the central concept toward better understanding the domain and the team and meeting patient care needs. Learning from experience is important; therefore, an idea offered was to *swap hats* to reflect on and understand roles, for example, by encouraging the engineer to take the patient's history and the clinician to be aware of the evolution of technology. With regard to teamwork research, a series of interesting ideas emerged. Discovering and highlighting the value of participating and working in an interprofessional or interdisciplinary team was identified as an essential process. To alleviate barriers, it is crucial to build interprofessional or interdisciplinary empathy skills as well as to understand and communicate one's own skills to strengthen the team. Each discipline may have a different view (engineers, computer scientists, and medical doctors); therefore, an iterative procedure is needed. Participants emphasized the importance of remaining open to new ideas, understanding the gaps, and highlighting both success and failure stories. The teamwork qualities mentioned most frequently were curiosity, attitude, confidence within the team, and acceptance of complementary personality types.

Discussion

Overview

Our research suggests two areas where improvements can be made in the synthesis and design of new CH education concepts: (1) overcoming disciplinary boundaries and (2) designing for disciplinary interaction. These build into a model for IDE for CH as illustrated in Figure 1.

Figure 1. A model for interdisciplinary education for connected health.



Overcoming Disciplinary Boundaries

Our research identifies three challenges that must be overcome in the design of new CH education programs to ensure cross-disciplinary engagement: communication challenges, state-of-the-art knowledge, and resistance to interdisciplinarity. We have combined insights from both the literature and the data gathered to develop suggested responses to the challenges identified as outlined below.

Challenge 1: Communication

In the first instance, our data indicated that difficulties regarding the different perceptions of ideas and concepts were hindering

interdisciplinary endeavors. Participants also highlighted the difficulties posed by multiple disciplinary *jargons*.

Proposed Solution

On the basis of participants' insistence that understanding alternate perspectives is key, we borrow from design thinking literature to suggest the use of empathy interviews to gain an understanding of multiple viewpoints. Sarasvarthy [47] presents it in the context of entrepreneurial thinking, "the point of exploring contrasting perspectives...is not to prove one superior to the other, but to learn to understand and use both". Addressing the disciplinary language difficulty, we combine interview and workshop data to suggest the investment of time at the master's level to develop a common language directed at CH activities that span disciplinary boundaries.

Challenge 2: State-of-the-Art Knowledge

The second major challenge raised throughout our research was the divergence in knowledge across disciplinary boundaries as to the availability of state-of-the-art technologies.

Proposed Solution

We accept participants' recommendations from our workshops that students be encouraged to *swap hats* throughout their training to develop an understanding of the different roles of different disciplines. This may increase understanding on all sides of the disciplinary divide that their interdisciplinary colleagues are only like to be at the cutting edge of their own discipline. Just as computer scientists are unaware of the latest medical techniques, clinicians are unlikely to know of the latest software developments. This will, in turn, lead to patience and a willingness to invest time to bring interdisciplinary colleagues up to speed.

Challenge 3: Resistance to Interdisciplinarity

Our data indicate an unwillingness among some students, researchers, and practitioners to reach across the disciplinary divide.

Proposed Solution

We recommend iterative processes in the education modules as per the principles of design-based learning discussed in the Introduction section [39,40]. Social constructivist learning theory, whereby learners work together, sharing their learning and constructing a new understanding for themselves through their experiences, underpins the concept of design-based learning [48]. Our data indicate that this is required to ensure that different views are accommodated and that interdisciplinary skeptics are given time and repeated opportunities to come on board.

Designing for Disciplinary Interaction

Our research indicates two major factors that lead to the successful design of interdisciplinary CH education: (1) the nature of the education module and (2) the positioning of the module within the education program.

The Nature of the Education Module

Overall, two key issues arose as to the nature of the education experience when seeking to integrate disciplines that may assist in the design of successful CH education going forward. The first is the creation of an enjoyable and challenging educational environment. Our data suggest that innovation and interest are key to the design of such an environment. The second is the necessity to link research and education together with its corresponding outcome, as our study confirms that education recipients wish to map their educational and research activities onto real-world practice. We recommend the use of both problem and project-based learning approaches to address both environment and application needs. Our workshop results suggest that although project-based activities are more suited to health- and well-being–focused challenges and educational outcomes, problem-based learning is more likely to address those that focus on patient care.

The Positioning of the Module Within the Education Program

Our data suggested a divergence of opinion as to whether these modules and the learning associated should be delivered early or late in the educational life cycle. Arguments for early delivery center on the advantages of early integration and the embedding of interdisciplinary attitudes. Those who advocated later delivery were concerned that the lack of developed skills and knowledge at that point might hinder useful interdisciplinarity. In response to both sets of concerns, we recommend a dual approach suggested by our workshop participants, which sees early awareness-raising activities (at the undergraduate level) followed by later problem- and project-based learning activities (at the master's level). Introducing interdisciplinary CH learning in this way is supported by Schön's work in relation to the role of reflection in professional learning discussed by Atkins and Murphy [49], who maintain that reflective professional learning occurs in three stages: first, creating awareness and feelings of discomfort; second, critically analyzing knowledge and feelings; and finally, developing a new perspective. In a similar vein, the spiral curriculum [50], an education design whereby topics are visited and revisited at increasing levels of complexity, allows the development of deep learning. Theory, therefore, supports our participants' suggestion of an early introduction coupled with a later development of learning.

Conclusions and Future Work

The CH ecosystem sets the basis for the investigation and deployment of new care models, leveraged by technology [51,52]. CH offers new opportunities for redesign and improvement of health and care, but its implementation and acceptance necessitate reorganization at multiple levels. IDE in CH is the cornerstone for broader adoption and impact of the CH paradigm and a prerequisite for research advancements in the CH field. However, such educational activities are not widely developed [11], serving as motivation for this study that focuses on identification and understanding of barriers and challenges in CH IDE.

This study relied on surveyed opinions and views across multiple disciplines, attempting to provide some insights on the existing challenges and to indicate potential directions for (successful) implementation of CH education. It mainly focused on the inherent interdisciplinarity in the field, explored the interdisciplinary-related barriers, and offered solutions for

overcoming interdisciplinary boundaries and designing CH curriculum.

We consider this work as a first step in investigating pathways to successful CH education. Although this work might be limited in terms of number of organizations involved in the workshops used to define the themes and challenges of interdisciplinary education, the number of identified disciplines mentioned in the obtained themes and challenges reassures the quality of achieved results. A wider multicountry mapping of needs and ideas would further contribute in terms of context-related barriers and enablers as well as in the formation of multicultural CH educational networks. To create solid evidence in this area, the work of Car [53], which proposes a methodology for systematic reviews in digital health education, can be expanded in this direction, encompassing the interdisciplinary aspects of CH education. In addition, recent evidence that new digital education tools such as virtual reality can improve knowledge and skills of health professionals [54] indicates that the challenges of interdisciplinary CH educations could be further explored via virtual reality and similar innovative means.

This work contributes useful inputs for CH curricula design with a focus on interdisciplinarity, both with regard to the alleviation of barriers and the design of interaction between different stakeholders in CH.

Acknowledgments

This article is based upon work from COST Action ENJECT TD 1405, supported by COST (European Cooperation in Science and Technology; www.cost.eu). The participants of the two workshops in Dublin and the interview participants in Thessaloniki are especially acknowledged.

Authors' Contributions

IC and TC designed the study. IC conducted the workshops at UCD, with the support of TC, and principally edited the manuscript. NM contributed toward the consolidation of results, synthesis of concepts, and editing of the manuscript. VT and TLT contributed with ideas for the CH education paper, literature enhancement, and editing of the manuscript. TC (corresponding author) coordinated the whole study, conducted the interviews at AUTH with the support of IC, and edited the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interview schedule.
[PDF File (Adobe PDF File), 45 KB - jmir_v21i11e14020_app1.pdf]

Multimedia Appendix 2
Workshop 1 questions.
[PDF File (Adobe PDF File), 71 KB - jmir_v21i11e14020_app2.pdf]

Multimedia Appendix 3
Workshop 2 questions.
[PDF File (Adobe PDF File), 46 KB - jmir_v21i11e14020_app3.pdf]

References

1. Deloitte. Global health care outlook URL: https://www2.deloitte.com/content/dam/Deloitte/au/Documents/life-sciences-health-care/au-lshc-global-health-care-outlook-2018-140219.pdf [accessed 2019-11-29]
2. OECD iLibrary. 2018. Health at a Glance: Europe 2018 : State of Health in the EU Cycle URL: https://doi.org/10.1787/health_glance_eur-2018-en [accessed 2019-10-29]
3. Caulfield B, Donnelly S. What is Connected Health and why will it change your practice? Quart J Med 2013 Aug;106(8):703-707. [doi: 10.1093/qjmed/hct114] [Medline: 23676416]
4. Argent R, Daly A, Caulfield B. Patient involvement with home-based exercise programs: can connected health interventions influence adherence? JMIR Mhealth Uhealth 2018 Mar 1;6(3):e47 [FREE Full text] [doi: 10.2196/mhealth.8518] [Medline: 29496655]
5. Hoogland J, Wijnen A, Munsterman T, Gerritsma CL, Dijkstra B, Zijlstra WP, et al. Feasibility and patient experience of a home-based rehabilitation program driven by a tablet app and mobility monitoring for patients after a total hip arthroplasty. JMIR Mhealth Uhealth 2019 Jan 31;7(1):e10342 [FREE Full text] [doi: 10.2196/10342] [Medline: 30702438]
6. Fernandez-Luque L, Staccini P. All that glitters is not gold: consumer health informatics and education in the era of social media and health apps. Yearb Med Inform 2016;25(1):188-193. [doi: 10.15265/iy-2016-045]

7. Loiselle CG, Ahmed S. Is connected health contributing to a healthier population? J Med Internet Res 2017 Nov 10;19(11):e386 [FREE Full text] [doi: 10.2196/jmir.8309] [Medline: 29127077]
8. Mountford N, Chouvarda I, Isomursu M, Caulfield B. Vectors and drivers of connected health in Europe: a foundation for integrated care. Int J Integr Care 2017;17(5):305 [FREE Full text] [doi: 10.5334/ijic.3622]
9. Gröne O, Garcia-Barbero M. Integrated care. Int J Integr Care 2001 Jun 01;1(2):-. [doi: 10.5334/ijic.28]
10. Xu L, Simjanoska M, Koteska B, Trajkovikj V, Bogdanova AM, Staric KD, et al. What clinics are expecting from data scientists? A review on data oriented studies through qualitative and quantitative approaches. IEEE Access 2019;7:641-654. [doi: 10.1109/ACCESS.2018.2885586]
11. Jonas CE, Durning SJ, Zebrowski C, Cimino F. An interdisciplinary, multi-institution telehealth course for third-year medical students. Acad Med 2019;94(6):833-837. [doi: 10.1097/acm.0000000000002701]
12. Pathipati AS, Azad TD, Jethwani K. Telemedical education: training digital natives in telemedicine. J Med Internet Res 2016 Jul 12;18(7):e193 [FREE Full text] [doi: 10.2196/jmir.5534] [Medline: 27405323]
13. Slovensky DJ, Malvey DM, Neigel AR. A model for mHealth skills training for clinicians: meeting the future now. mHealth 2017;3:24 [FREE Full text] [doi: 10.21037/mhealth.2017.05.03] [Medline: 28736733]
14. Saranto K, Ronquillo C, Velez O. Nursing competencies for multiple modalities of connected health technologies. Stud Health Technol Inform 2017;232:172-182. [Medline: 28106595]
15. Mountford N, Zubiete ED, Kessie T, Garcia-Zapirain B, Nuño-Solinís R, Coyle D, et al. Activating technology for connected health in cancer: protocol for a research and training program. JMIR Res Protoc 2018 Jan 24;7(1):e14 [FREE Full text] [doi: 10.2196/resprot.8900] [Medline: 29367184]
16. COST | European Cooperation in Science and Technology. ENJECT COST Action: European Network for the Joint Evaluation of Connected Health Technologies URL: https://www.cost.eu/actions/TD1405/#tabs|Name:overview [accessed 2019-10-16]
17. Mountford N, Kessie T, Watts G, Fernandez L, Chouvarda I, Cusack T. An Interdisciplinary 4th Level Education Model: Connected Health. In: Proceedings of the 3rd International Conference on Higher Education Advances. 2017 Presented at: HEAd'17; June 21-23, 2017; Valencia, Spain URL: https://researchrepository.ucd.ie/handle/10197/8517 [doi: 10.4995/head17.2017.5485]
18. Oandasan I, Reeves S. Key elements for interprofessional education. Part 1: the learner, the educator and the learning context. J Interprof Care 2005 May;19(Suppl 1):21-38. [doi: 10.1080/13561820500083550] [Medline: 16096143]
19. Barr H, Gray R, Helme M, Low H, Reeves S. Interprofessional Education Guidelines 2016. UK: CAIPE; 2016.
20. L'Ecuyer KM, Pole D, Leander SA. The use of PBL in an interprofessional education course for health care professional students. Interdiscip J Probl Based Learn 2015;9(1):- [FREE Full text] [doi: 10.7771/1541-5015.1497]
21. Woods C. Researching and developing interdisciplinary teaching: towards a conceptual framework for classroom communication. High Educ 2007;54(6):853-866. [doi: 10.1007/s10734-006-9027-3]
22. Smith S, Clouder L. Interprofessional and interdisciplinary learning: an exploration of similarities and differences. In: Bromage A, Clouder L, Thistlethwaite J, Gordon F, editors. Interprofessional E-Learning and Collaborative Work: Practices and Technologies. Hershey, PA: IGI Global; 2010:1-13.
23. Aboelela SW, Larson E, Bakken S, Carrasquillo O, Formicola A, Glied SA, et al. Defining interdisciplinary research: conclusions from a critical review of the literature. Health Serv Res 2007 Feb;42(1 Pt 1):329-346 [FREE Full text] [doi: 10.1111/j.1475-6773.2006.00621.x] [Medline: 17355595]
24. Hall P. Interprofessional teamwork: professional cultures as barriers. J Interprof Care 2005 May;19(Suppl 1):188-196. [doi: 10.1080/13561820500081745] [Medline: 16096155]
25. Spoelstra H, Stoyanov S, Burgoyne L, Bennett D, Sweeney C, Drachsler H, et al. Convergence and translation: attitudes to inter-professional learning and teaching of creative problem-solving among medical and engineering students and staff. BMC Med Educ 2014 Jan 22;14:14 [FREE Full text] [doi: 10.1186/1472-6920-14-14] [Medline: 24450310]
26. Feyerabend P. How to defend society against science. In: Klemke E, Hollinger R, Rudge D, Kline AD, editors. Introductory Readings in the Philosophy of Science.Third Edition. Amherst, NY: Prometheus; 1998:54-65.
27. Mountford N, Coleman M, Kessie T, Cusack T. Interdisciplinary doctoral research networks: enhancers and inhibitors of social capital development. Stud High Educ 2019:1-16 [FREE Full text] [doi: 10.1080/03075079.2019.1623768]
28. Cassi L, Champeimont R, Mescheba W, de Turckheim E. Analysing institutions interdisciplinarity by extensive use of Rao-Stirling diversity index. PLoS One 2017;12(1):e0170296 [FREE Full text] [doi: 10.1371/journal.pone.0170296] [Medline: 28114382]
29. Bromham L, Dinnage R, Hua X. Interdisciplinary research has consistently lower funding success. Nature 2016 Jun 30;534(7609):684-687. [doi: 10.1038/nature18315] [Medline: 27357795]
30. Larivière V, Haustein S, Börner K. Long-distance interdisciplinarity leads to higher scientific impact. PLoS One 2015;10(3):e0122565 [FREE Full text] [doi: 10.1371/journal.pone.0122565] [Medline: 25822658]
31. Yegros-Yegros A, Rafols I, D'Este P. Does interdisciplinary research lead to higher citation impact? The different effect of proximal and distal interdisciplinarity. PLoS One 2015;10(8):e0135095 [FREE Full text] [doi: 10.1371/journal.pone.0135095] [Medline: 26266805]

32. Stirling A. Semantic Scholar. 1998. On the economics and analysis of diversity URL: https://pdfs.semanticscholar.org/08e8/fb40fdbfd91fec63fc632d91b7dbc5779ea1.pdf [accessed 2019-11-01]
33. Stirling A. A general framework for analysing diversity in science, technology and society. J R Soc Interface 2007 Aug 22;4(15):707-719 [FREE Full text] [doi: 10.1098/rsif.2007.0213] [Medline: 17327202]
34. Voorhees RA. CiteSeerX. 2001. Competency-Based Learning Models: A Necessary Future URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.502.8016&rep=rep1&type=pdf [accessed 2019-11-01]
35. Hall Barber K, Schultz K, Scott A, Pollock E, Kotecha J, Martin D. Teaching quality improvement in graduate medical education. Acad Med 2015;90(10):1363-1367. [doi: 10.1097/acm.0000000000000851]
36. Cusack T, O'Donoghue G, Butler M, Blake C, O'Sullivan C, Smith K, et al. A pilot study to evaluate the introduction of an interprofessional problem-based learning module. Interdiscip J Probl Based Learn 2012;6(2):31 [FREE Full text] [doi: 10.7771/1541-5015.1350]
37. Helle L, Tynjälä P, Olkinuora E. Project-based learning in post-secondary education – theory, practice and rubber sling shots. High Educ 2006;51(2):287-314. [doi: 10.1007/s10734-004-6386-5]
38. Dolmans DH, Tigelaar D. Building bridges between theory and practice in medical education using a design-based research approach: AMEE Guide No. 60. Med Teach 2012;34(1):1-10. [doi: 10.3109/0142159X.2011.595437] [Medline: 22250671]
39. Peffer M, Renken M. Practical strategies for collaboration across discipline-based education research and the learning sciences. CBE Life Sci Educ 2016;15(4):pii: es11 [FREE Full text] [doi: 10.1187/cbe.15-12-0252] [Medline: 27881446]
40. Stokholm M. Problem based learning versus design thinking in team based project work. In: Kovacevic A, Eger A, Parkinson BE, editors. Design Education & Human Technology Relations. UK: Institution of Engineering Designers, The Design Society; 2014:268-274.
41. Brennan L, Cusack T, Delahunt E, Kuznesof S, Donnelly S. Academics' conceptualisations of the research-teaching nexus in a research-intensive Irish university: A dynamic framework for growth & development. Learn Instruct 2019;60:301-309. [doi: 10.1016/j.learninstruc.2017.10.005]
42. McMahon S, O'Donoghue G, Doody C, O'Neill G, Barrett T, Cusack T. Standing on the precipice: evaluating final-year physiotherapy students' perspectives of their curriculum as preparation for primary health care practice. Physiother Can 2016;68(2):188-196. [doi: 10.3138/ptc.2015-11e]
43. Galvin R, Stokes E, Cusack T. Family-Mediated Exercises (FAME): an exploration of participant's involvement in a novel form of exercise delivery after stroke. Top Stroke Rehabil 2014;21(1):63-74. [doi: 10.1310/tsr2101-63] [Medline: 24521841]
44. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006;3(2):77-101 [FREE Full text] [doi: 10.1191/1478088706qp063oa]
45. Joffe H. Thematic analysis. In: Harper D, Thompson AR, editors. Qualitative research methods in mental health and psychotherapy. In Qualitative Research Methods in Mental Health and Psychotherapy: A Guide for Students and Practitioners. Edited by David Harper and Andrew Thompson, Chichester: Wiley-Blackwell. USA: Wiley; 2012:209-223.
46. O'Donoghue G, Doody C, Cusack T. Using student centred evaluation for curriculum enhancement: An examination of undergraduate physiotherapy education in relation to physical activity and exercise prescription. Stud Educ Eval 2011;37(2-3):170-176 [FREE Full text] [doi: 10.1016/j.stueduc.2011.04.004]
47. Sarasvathy SD. Effectuation: Elements of Entrepreneurial Expertise. UK: Edward Elgar Publishing; 2009.
48. Driscoll M. Psychology of Learning for Instruction. Second Edition. Boston: Allyn & Bacon; 1999.
49. Atkins S, Murphy K. Reflection: a review of the literature. J Adv Nurs 1993 Aug;18(8):1188-1192. [doi: 10.1046/j.1365-2648.1993.18081188.x] [Medline: 8376656]
50. Harden RM. What is a spiral curriculum? Med Teach 1999;21(2):141-143. [doi: 10.1080/01421599979752] [Medline: 21275727]
51. Chouvarda IG, Goulis DG, Lambrinoudaki I, Maglaveras N. Connected health and integrated care: Toward new models for chronic disease management. Maturitas 2015 Sep;82(1):22-27. [doi: 10.1016/j.maturitas.2015.03.015] [Medline: 25891502]
52. Chen SC. Technological health intervention in population aging to assist people to work smarter not harder: qualitative study. J Med Internet Res 2018 Jan 4;20(1):e3 [FREE Full text] [doi: 10.2196/jmir.8977] [Medline: 29301736]
53. Car J, Carlstedt-Duke J, Tudor Car L, Posadzki P, Whiting P, Zary N, Digital Health Education Collaboration. Digital education in health professions: the need for overarching evidence synthesis. J Med Internet Res 2019 Feb 14;21(2):e12913 [FREE Full text] [doi: 10.2196/12913] [Medline: 30762583]
54. Kyaw BM, Saxena N, Posadzki P, Vseteckova J, Nikolaou CK, George PP, et al. Virtual reality for health professions education: systematic review and meta-analysis by the digital health education collaboration. J Med Internet Res 2019 Jan 22;21(1):e12959. [doi: 10.2196/12959] [Medline: 30668519]

Abbreviations

AUTH: Aristotle University of Thessaloniki
CH: connected health
COST: Cooperation in Science and Technology

ENJECT: European Network for the Joint Evaluation of Connected Health Technologies
EU: European Union
ICT: information and communications technology
IDE: interdisciplinary education
IPE: interprofessional education
mHealth: mobile health
MI-LAB: Lab of Computing and Medical Informatics and Biomedical Imaging Technologies
PBL: problem-based learning
UCD: University College Dublin

Edited by M Tilney; submitted 15.03.19; peer-reviewed by J Browne, S Chen; comments to author 24.06.19; revised version received 15.09.19; accepted 22.10.19; published 13.11.19.

Please cite as:
Chouvarda I, Mountford N, Trajkovik V, Loncar-Turukalo T, Cusack T
Leveraging Interdisciplinary Education Toward Securing the Future of Connected Health Research in Europe: Qualitative Study
J Med Internet Res 2019;21(11):e14020
URL: http://www.jmir.org/2019/11/e14020/
doi:10.2196/14020
PMID:31719026

Original Paper

Connected Health User Willingness to Share Personal Health Data: Questionnaire Study

Maria Karampela[1], MSc; Sofia Ouhbi[2], PhD; Minna Isomursu[3], PhD

[1]IT University of Copenhagen, Copenhagen S, Denmark

[2]United Arab Emirates University, Al Ain, United Arab Emirates

[3]University of Oulu, Oulu, Finland

Corresponding Author:
Maria Karampela, MSc
IT University of Copenhagen
Rued Langgaards Vej 7
Copenhagen S,
Denmark
Phone: 45 25 54 09 84
Email: makar@itu.dk

Abstract

Background: Connected health has created opportunities for leveraging health data to deliver preventive and personalized health care services. The increasing number of personal devices and advances in measurement technologies contribute to an exponential growth in digital health data. The practices for sharing data across the health ecosystem are evolving as there are more opportunities for using such data to deliver responsive health services.

Objective: The objective of this study was to explore user attitudes toward sharing personal health data (PHD). The study was executed within the first year after the implementation of the new General Data Protection Regulation (GDPR) legal framework.

Methods: The authors analyzed the results of an online questionnaire survey to explore the willingness of 8004 people using connected health services across four European countries to share their PHD and the conditions under which they would be willing to do so.

Results: Our findings indicate that the majority of users are willing to share their personal PHD for scientific research (1811/8004, 22.63%). Age, education level, and occupation of the participants, in addition to the level of digitalization in their country were found to be associated with data sharing attitudes.

Conclusions: Positive attitudes toward data sharing for scientific research can be perceived as an indication of trust established between users and academia. Nevertheless, the interpretation of data sharing attitudes is a complex process, related to and influenced by various factors.

(J Med Internet Res 2019;21(11):e14537) doi:10.2196/14537

KEYWORDS

connected health; personal health data; data sharing; questionnaire

Introduction

Background

The evolution of information and communication technologies in health care holds promise for the reformation of health care services. The development of novel technologies and their acceptance by health care providers has an impact on the design and delivery of services, resulting not only in a greater number but also better quality of services [1]. Similarly, the proliferation of sensors, either worn or in mobile devices and living spaces, contributes to the creation of a new landscape for health care services, leading the way for tailor-made interventions [2,3]. According to the International Data Corporation, 1.42 billion smartphones were sold globally in 2018 [4], and by 2020, the number of smartphone users is projected to reach 2.1 billion [5], while the number of sensors connected to the Internet of Things (IoT) is forecasted to reach between 30 and 50 billion [6]. In light of these facts, the adoption of new technologies has resulted in an increasing amount of personal data stored across various databases [7,8]. The potential benefits of using big data in health care have not only increased user expectations of more sophisticated and personalized solutions but also accelerated

health care professional and practitioner efforts toward the creation of more agile connected health systems to communicate health data [9,10]. Connected health can be defined as "the collective term" regrouping "telecare, telehealth, telemedicine, mHealth, digital health, and eHealth services" [11]. It is "a comprehensive, sociotechnical model for managing health care through software solutions" [12]. Moreover, it is transforming health care services [13,14], connecting patients with caregivers and clinicians [15], and empowering them to assume greater responsibility for their own health care decisions [16]. Connected health provides individual patients, patients with chronic conditions, and users in traditionally underserved populations the potential for enhanced health care service reach at a relatively low cost; it also improves scalability, time efficiency, and tailoring and customization [17,18].

The use of personal health data (PHD) can be essential to the evolution of health care systems in general [19] and connected health systems in particular [20]. Sharing health data via connected health solutions can have a positive social impact by promoting social solidarity among patients, a positive economic impact by optimizing health care resources, and a positive environmental impact by reducing the need for transportation [21]. Connected health solutions for disease prevention, like Flu Near You [22] or HealthMap [23] that map flu symptoms and infectious disease outbreaks, are examples of the potential for commercial data sharing to prevent contagious diseases. From the perspective of patients with chronic conditions, sharing health data can be valuable in many ways. For instance, it can be used to obtain a better understanding of diagnosis and treatment options or better self-management of diseases [24,25]. Similarly, for health care professionals, potential uses are numerous, as health data can facilitate the exploring of new therapeutic treatments for life-threatening diseases like cancer [26]. Data sharing initiatives have also reinforced legal frameworks and contributed to the promotion of more transparent regulations [27-29]. From a technical point of view, the need for using health data has led to the development of interoperable connected health systems for secure data exchange [30]. Frameworks to improve privacy in cloud computing and encryption techniques to offer secure access control have been among the proposed solutions to providing reliable data sharing interventions [31,32]. Despite the potential benefits of data utilization, user willingness to share personal data is changing [33]. Therefore, understanding user attitudes is crucial for developing connected health systems in the future.

The literature that examines user behavior in terms of technology adoption and use draws on various theoretical perspectives—for example, the technology acceptance model or theory of reasoned action [32]. These theories support the arguments that users develop attitudes toward technology and there is a positive relationship between attitudes and behavioral control [32,34]. Several studies have been focusing on expanding this relationship in the context of technology [32,35]. In the context of our study, user intentions to perform actions "are assumed to capture the motivational factors that influence a behavior; they are indications of how hard people are willing to try, and of how much of an effort they are planning to exert, in order to perform the behavior" [34]. Based on this definition, a willingness to share PHD can be indicative of connected health user intention to perform an action in a given situation. In this study, the conditions under which connected health users would be willing to share their PHD were limited to a set of specific situations; thus, user willingness to share their data will be discussed within this context.

Digital technology user attitudes toward sharing different types of personal data have been examined in previous studies [33]. A study by Athey et al [36] concluded that users are willing to share personal information such as private emails in exchange for small incentives such as pizza. Pickard and Swan [37] argued that consumers, in exchange of financial rewards, would be willing to sell their anonymized health data [37]. A study by Chen et al [38] about wellness data sharing concluded that user data sharing willingness is dependent on the potential uses of the data. Interestingly, this notion is further supported when considering services like 23andMe [39] for DNA analysis and PatientsLikeMe [40] for networking. Nevertheless, willingness to share health data was found to be dependent on various parameters such as the perceived sensitivity of the information [41]. While the sharing of other types of data, related to consumption and finances for example, has been seen to entail some privacy concerns, the sharing of health data is far more complicated [42,43].

Willingness to share PHD has been found to be dependent on various other parameters. For example, users are motivated to share their PHD in exchange for care improvements, better public health [44], or health information exchange [45]. Pickard and Swan's study [37] concluded that attitudes toward data sharing for scientific purposes are becoming more positive. Additionally, users make a distinction between deidentified data and data with personally identifiable information. In a study by Weng et al [46], the majority of the respondents (89%) stated that sharing anonymized clinical data for research is preferable to sharing identifiable data. These studies confirm the idea that users desire to have control over their data sharing preferences. Individual control over data sharing preferences and patient-centered sharing models have been indicated to have positive effects on user data sharing willingness [47,48]. For example, 59% of the participants in a study by Weitzman et al [49] stated that they would prefer an opt-in data sharing model.

Apart from control over data sharing preferences, knowledge and trust have been found to contribute to more positive attitudes. Previous findings support the argument that it is unimportant for participants to fully comprehend different types of consent; nevertheless an opt-in model of informed consent is valued as a more trusted data sharing practice [50,51]. The relationship between consent and trust is fragile. Providing consent to share data cannot be equated to having trust [52]. When individuals consider sharing health information, they often weigh the personal benefits against potential risks [53-55]. Concerns of health information privacy are found to be a common barrier in sharing [56-60]. In this study, user willingness to share PHD is concerned with attitudes toward sharing health data. Data that are available online are indicative of different facets of people's lives and can be informative about their health [8,61]. These digital traces of everyday life are the product of users' active or passive interaction with the network

[62]. Therefore, in the context of this study, PHD refers to the available online information that is the product of users' active or passive interaction with the network and is indicative of their health. Data sharing attitudes have been explored in previous research. Nevertheless, the implementation of the new General Data Protection Regulation (GDPR) less than a year ago has led to renewed interest in this topic.

Objectives

Over the years, researchers have been paying attention to people's willingness to share their PHD. As technology and legal frameworks evolve, user attitudes are also undergoing a shift. This study explores and discusses current attitudes across four European countries and addresses the following research questions (RQs):

> *RQ1: Are connected health users willing to share their PHD?*
>
> *RQ2: Under what conditions are they are willing to share their PHD?*

In order to interpret the results of the study, the RQs will be discussed along the following parameters: participant age, area of residence, country of residence, education, and occupational group.

Methods

Research Design

The purpose of this study was to investigate user willingness and conditions for sharing their PHD. This paper presents findings from a household questionnaire survey designed by Sitra Innovation Fund and distributed by Kantar TNS Oy, a global market research and information group, in December 2018. Sitra is a Finnish Innovation Fund that through its research aims to influence European policy makers toward more sustainable well-being on social, financial, and ecological levels. The data used in this study are a subset of data from large-scale research conducted within the framework of Sitra's IHAN project [63]. The aim of the IHAN project is twofold: to develop foundations for a fair and human-driven data economy by creating a method for data exchange and influence regulatory development toward fair use of data through European Union (EU) policy makers.

In regard to the research design of the survey, author MI has been actively involved in its development as a scientific advisor. The content of the survey was designed by Sitra through collaborative team processes to get an overall understanding of the data-driven attitudes after the implementation of the new GDPR. The website survey design was developed by Kantar TNS Oy in collaboration with Sitra. Quantitative surveys have been used by prior studies to explore and discuss user perspectives [64-66]. A survey questionnaire was chosen as research design because our goal was to get an understanding of how European countries are in regard to data economy–related topics. More specifically, we aimed to present an overview of the current landscape pertinent to user willingness and conditions for sharing their PHD after the implementation of the new GDPR. Aligned with this aim, the survey presented here was one of the first steps in a multiyear pilot project administered by Sitra. A quantitative method was chosen by Sitra to give an overview of data economy–related topics from the citizen's point of view, and later steps of the IHAN project will investigate the findings highlighted by the survey with more qualitative approaches.

Survey Questions

The original online survey comprised 27 questions, both open- and close-ended, to collect data relevant to background characteristics of respondents, attitudes toward services, and trust toward services and data management. For the purpose of this paper, none of the open-ended questions were used, and we only considered two of the closed-ended questions. The questions we chose to include were close-ended, as the goal was to get a quantitative overview of attitudes. The two questions measured user willingness to share and conditions of sharing their PHD.

> *RQ1: Are connected health users willing to share their PHD?*
>
> *RQ2: Under what conditions are they are willing to share their PHD?*

Data sharing in this context was pertinent to various purposes such as scientific research, public interest, and in exchange for services or financial benefits. Additionally, several sociodemographic background questions were administered to the participants to capture factors that might influence user willingness. The questionnaire was translated by professional translators from Finnish to the official language of each country. For the purpose of analysis and reporting, the questions and related responses were translated back to Finnish and English to ensure a common understanding. The survey was conducted anonymously in compliance with the EU GDPR legal framework.

Data Collection

Four European countries were chosen, as GDPR is an EU regulation, even though it has global implications. Data collection was carried out from December 6 to 18, 2018, in Finland, the Netherlands, Germany, and France. The inclusion criteria were consent for participation in the survey and a self-declaration of being at least 18 years of age. Random sampling was performed by Kantar TNC Oy and was representative of the age, gender, and locality. Participants were invited to participate in the survey via phone or email communication. The questionnaire was delivered in the official language of each country. All participants were given financial compensation upon the completion of the online survey (Multimedia Appendix 1). The average completion time of the online survey was 12 minutes. Web-based sampling has been seen to be an effective medium to recruit participants, as it enables remote access and complies with individual preferences [67]. To comply with the emerging requirements of quality data collection in online surveys, we report our results based on the Checklist for Reporting Results of Internet E-Surveys (CHERRIES) [68].

Data Analysis

Analysis of the raw dataset was performed by Kantar TNS Oy using SPSS Statistics (IBM Corp) software. In this study, we used the descriptive statistics that we were given by Sitra and Kantar TNS Oy to generate visualizations. The data visualizations were used for the interpretation of the results.

Results

Overview

The online survey included a total of 8004 respondents: 2000 from Finland, 2004 from Germany, 2000 from the Netherlands, and 2000 from France. The sample was representative in terms of gender, age, and locality. The completion rates were 84.32% (2000/2372) in Finland, 53.27% (2004/3762) in Germany, 48.77% (2000/4101) in the Netherlands, and 54.10% (2000/3697) in France.

Demographics

Table 1 presents an overview of the participant demographics (values show the average of the percentage of all four countries). Among the participants, 49.00% (3922/8004) were male, 50.00% (4002/8004) were female, and 1.00% (80/8004) did not indicate their gender. The age distribution of the participants was between 18 and 65 years. Nearly a quarter (2001/8004, 25.00%) of the participants received compulsory education, 14.01% (1121/8004) received academic education, 58.00% (4642/8004) received other types of education, and 3.00% (240/8004) did not specify their education. Participants were also asked to specify the occupational group to which they belong, and 17.00% (1361/8004) were in a managerial position, 10.99% (880/8004) were in junior positions, 27.01% (2162/8004) were workers, 6.00% (480/8004) were self-employed or sole traders, 6.00% (480/8004) were students, 11.99% (960/8004) were pensioners, and 3.00% (240/8004) replied “don’t know.” The remaining 18.00% (1441/8004) of the participants did not fall within one of the previous categories. As for the types of living quarters, 40.00% (3202/8004) of them lived in a city, 34.00% (2721/8004) in a town or an urban area, 22.00% (1761/8004) in the countryside, and 4.00% (320/8004) did not know. See Multimedia Appendix 2 for the questionnaire survey results.

Table 1. Background information of respondents (N=8004).

Characteristics	Value[a], n (%)
Gender	
Male	3922 (49.00)
Female	4002 (50.00)
Other	80 (1.00)
Age in years	
18-34	2561 (32.00)
35-44	1521 (19.00)
45-65	3922 (49.00)
Region type	
City	3202 (40.00)
Town/urban area	2721 (34.00)
Countryside	1761 (22.00)
Do not know	320 (4.00)
Education	
Compulsory education	2001 (25.00)
Academic education	1121 (14.01)
Other[b]	4642 (58.00)
Do not know	240 (3.00)
Occupational group or status	
At school or student	480 (6.00)
Worker	2162 (27.01)
Self-employed or sole trader	480 (6.00)
Junior white collar	880 (10.99)
Managerial position/senior	1361 (17.00)
Pensioner	960 (11.99)
Other[c]	1441 (18.00)
Do not know	240 (3.00)

[a]Average of the percentage for all four countries.

[b]Other education corresponds to vocational education, matriculation, or other types of education.

[c]Other occupation corresponds to other types of jobs or status such as at-home mother/father.

Research Question Results

The majority of respondents were willing to share their PHD under specific conditions as shown in Table 2. The results differed according to the country the respondents were from. Figure 1 shows the country-wise distribution of the participant responses. Gender of the participants did not impact the results, although men (2234/3922, 56.96%) were slightly more willing to share their health data compared with women (2239/4002, 55.95%). Regarding age, young people were more willing to share their data than older people, as shown in Figure 2. Figure 3 shows that participants living in cities and urban areas were more willing to share their PHD compared with those living in the countryside. Figure 4 presents the results per education level of participants. Figure 5 presents the results according to the respondent occupation type.

Table 2. Participant responses about the conditions of sharing their personal health data (N=8004).

Responses	Value[a], n (%)
No	2384 (29.78)
Information is used for scientific research	1811 (22.63)
I would be paid for it	1139 (14.23)
I don't know	1092 (13.64)
Data is used for purposes of public interest	949 (11.86)
I would be offered extra services or individual service	628 (7.85)

[a]Average of the percentage for all four countries.

Figure 1. Results per country.

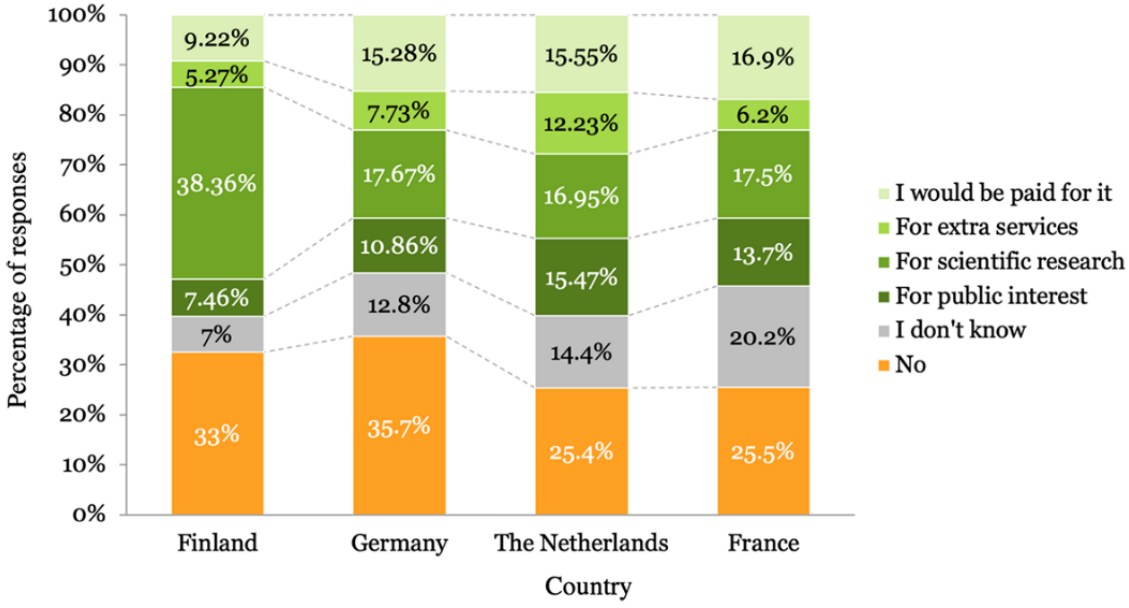


Figure 2. Results per age category.

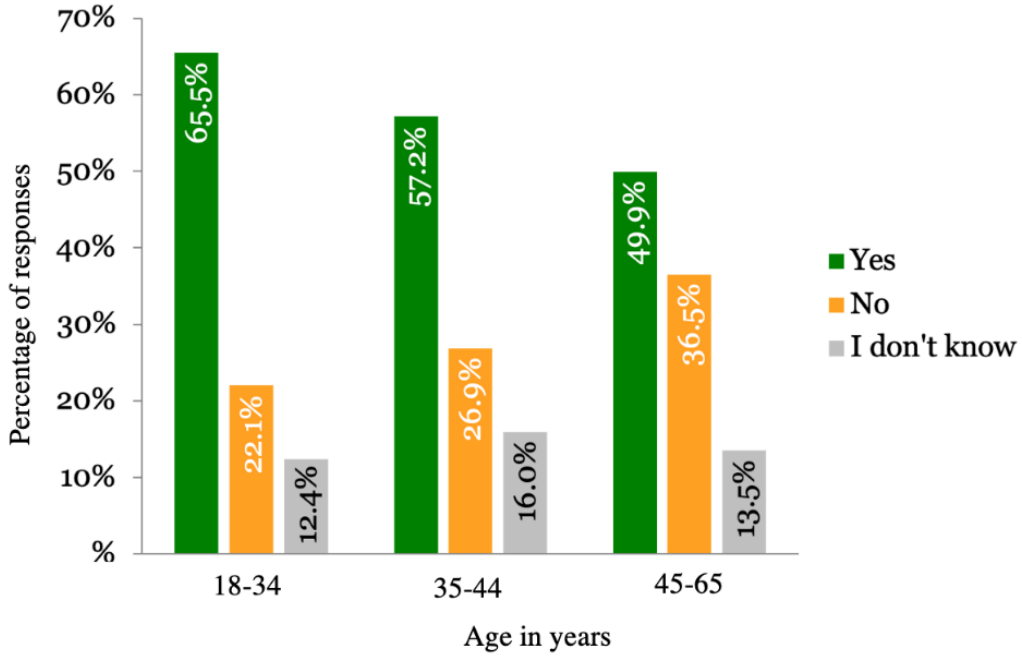


Figure 3. Results per region type.

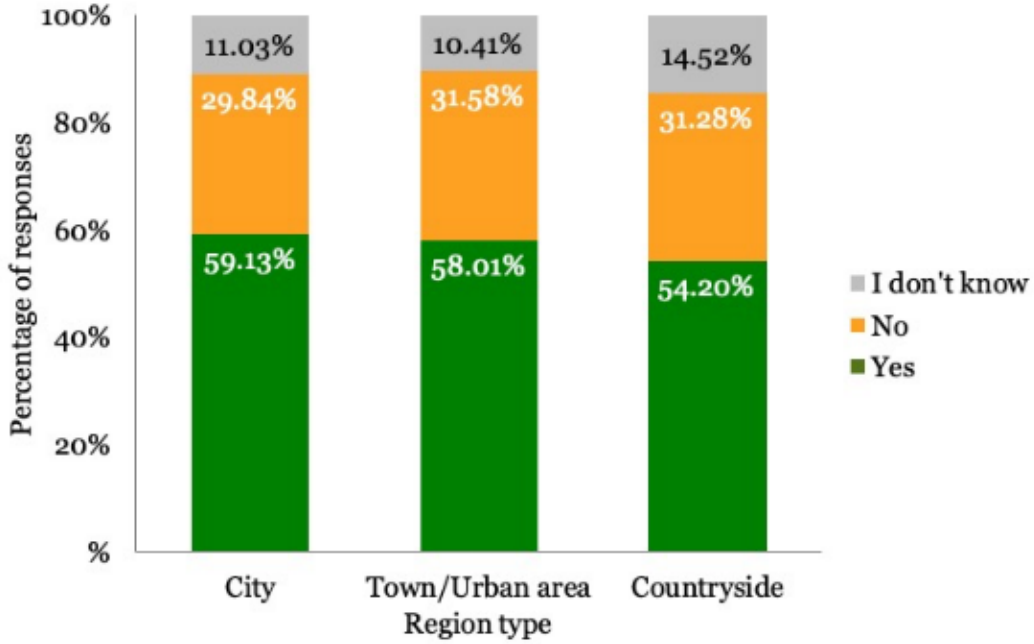


Figure 4. Results per education level.

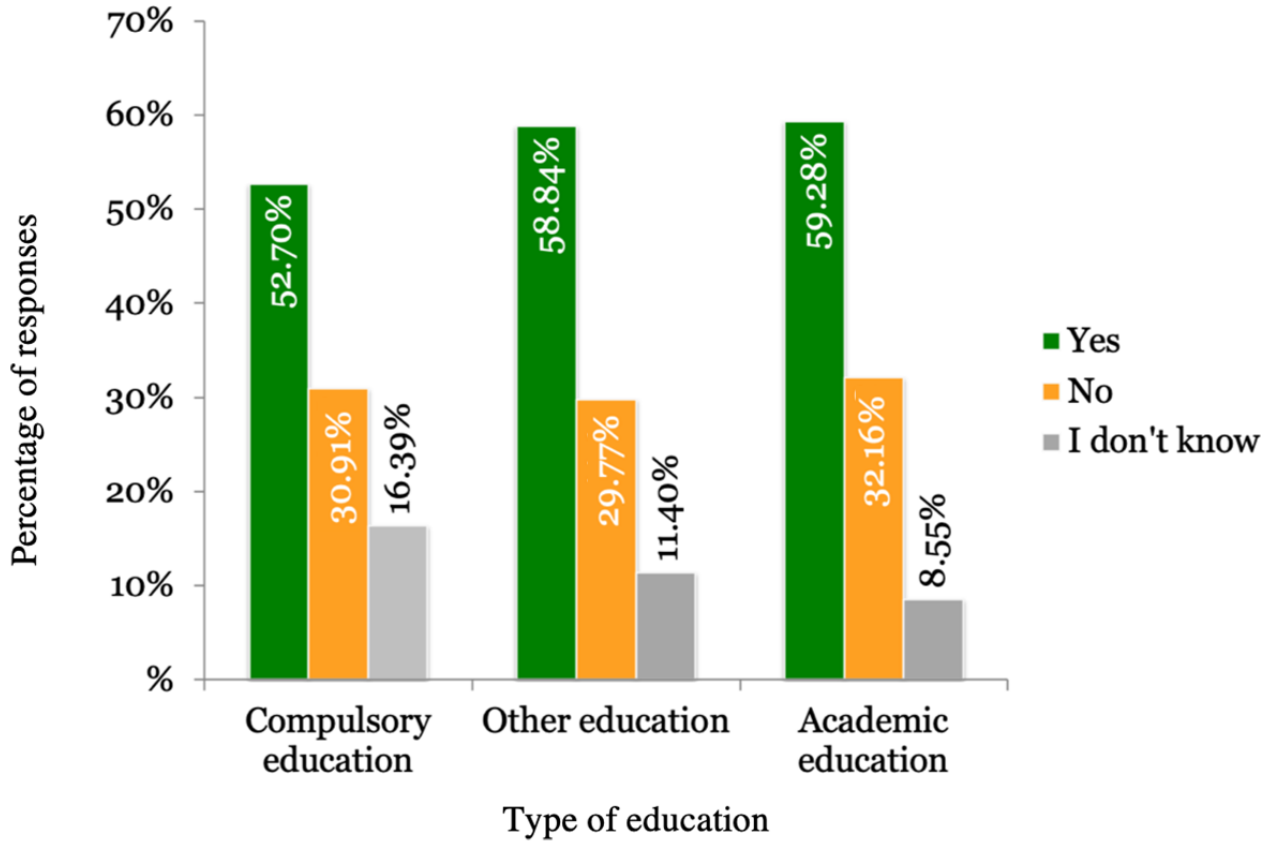


Figure 5. Results per occupational position type.

Discussion

Principal Findings

The aim of this study was to explore user willingness to share their digital PHD across four European countries. The four countries were chosen to represent countries in the EU that have a medium or high level of digitalization and availability of digital health data. Our findings show that the majority of users are willing to share their health data (RQ1), most often preferring to share it for scientific research (1811/8004, 22.63%; RQ2). User willingness was not strongly gender-related; however, our findings suggest that age, education, and occupation were prevalent factors affecting user attitudes.

In line with previous literature, our findings indicate that users are willing to share their PHD for research purposes [37,69]. This study extends previous research by providing deeper insights into the factors that shape user attitudes. The factors found to be affecting willingness were age, educational attainment, and occupational status. We found a negative relation between age and willingness; as users grow older, they tend to be more reluctant about data sharing, with pensioners having the highest percentage of negative responses (376/960, 39.17%). Older age has been connected with a lack of knowledge over factual risks and growing concerns over possible data misuse and security breaches [53,70]. Aging has been associated with the development of health conditions. Taking that into consideration along with the aversion of older people to sharing their health data due to privacy concerns can further support our results. Projections about the world population predict that in the future the older population will surpass younger. This change can introduce further implications in regards to data-sharing reluctance. Age has also been associated with digital literacy and competency in using digital devices [71,72]. Digital competency influences the way users interact with services [73] and is age-dependent in the sense that users grow up accustomed to different technologies. This means that they have different mental models regarding technology and the way it works [74,75]. In order to keep up with digitalization trends, users need to develop new competences or evolve their existing competences [76]. In regard to digital literacy, user willingness to exchange and use digital health information also decreases as they grow older, and the same trend applies to the willingness to learn to use online health technologies [71]. Education and occupation are also factors that influence user willingness. Lower levels of education are associated with lesser willingness to share data, possibly due to higher exposure to media and the misuse of information [54]. As for the occupational groups, pensioners were found to be the most reluctant to share PHD. Considering the similarity of trends in education and occupation, we could argue that higher educational attainment often leads to occupation at top-level or management positions, which has a positive impact on willingness to share PHD.

Current user trends in the four European countries were also informative about the impact of digitalization of health care services. The four countries that participated in the online survey belong to different groups in relation to the use of internet services. Finland and the Netherlands have the most advanced digital economies, while Germany and France belong to the medium performing group, with France ranking in the lowest position [77]. As for the level of digitalization in health care in EU, reports suggest that eHealth is still in its infancy. According to the Digital Economy and Society Index for Digital Public Services, less than 1 in 5 EU citizens have used eHealth services in the previous year [77]. Based on our results, users from countries that have medium performing digital economies were found to be more unwilling or unprepared to share their PHD. Compared with other European users, German users were the most reluctant to share their PHD, while more French users did not know if they would be willing to do so. In contrast, the Finnish participants, despite having the second highest number of negative responses toward willingness to share health data, had the highest percentage of willingness to share data for scientific research. The Dutch participants were the most willing to share their data. These findings likely correspond to the advanced levels of digitalization in Finnish and Dutch economies. However, culture-specific differences between the four countries could also play a role in the interpretation of the results. For instance, it has been reported in the past that German

are apprehensive about experiencing negative consequences as result of privacy violations [72].

In regard to the conditions under which users would be willing to share their PHD, scientific research and financial incentives were prevalent. This is consistent with previous findings [45,72,77]. Finnish users had the most positive attitudes toward sharing for scientific research, which is not surprising as Finland has the highest percentage of users of eHealth services [77], while in terms of the state of open government data publication, Finland ranks fifth and France fourth [78]. Reflecting upon this, we could argue that in both countries data are valued as fuel for research. France has a long history of strict data protection [79], legal flaws, vagueness, and data mishandling, as well as political instability, which might influence citizen attitudes toward data sharing [80]. Dutch users reported the highest willingness to share PHD for public interest. The case of the Netherlands is particular due to the fact that since 2012, the country has imposed stricter legislation for ePrivacy, requiring website users to have active control over their personal information online [81]. This may explain positive user attitudes toward public service providers.

While sharing for research purposes was the most prevalent condition, a disregard of public entities may be perceived as an indicator of trust issues. Incidents of data breaches in public health care are numerous [82], which may have a negative impact on users' data sharing attitudes. In contrast to Pickard and Swan's [37] study, many of our users responded that personalization of services is not a strong incentive for them to share their PHD [83]. Consistent with previous studies [44], older participants where less motivated to share their health data in exchange for financial benefits. As for education and occupation, those with higher qualifications preferred to share data for research purposes. In other words, higher educational attainment has a positive relation with willingness to share for scientific research. Although we do not have information about the health status of participants, an interesting aspect to consider is that patients with chronic or terminal diseases have more positive attitudes toward data sharing for scientific research [37,46,84]. Sharing data for research purposes has been related to altruism. From a cognitive standpoint, emotions may also play an essential role in data sharing willingness. In general, the role of emotions in decision making, specifically trust and concern, has been seen to have an effect on user behaviors [54].

Willingness toward sharing PHD is related with privacy and security considerations. The privacy calculus theory suggests that risk factors, such as the purpose for which the shared information will be used, are related to user willingness [55]. According to the theory, understanding user attitudes is a complicated task. Some researchers have claimed that stricter controls in privacy settings would increase user willingness [85], while others argue that factors such as social influence can lead to similar results [86]. Deidentification of data and an opt-in model of informed consent are also solutions that have been seen to have positive effects on data sharing attitudes [46,50,87]. Recent studies on secondary use of data have indicated that users are generally willing to accept the sharing of their health data for research purposes, without explicit consent [88]. User control over accessing and sharing PHD is reflected throughout previous research for centralizing patient data, and, therefore, future interoperable health care interventions should be aware of these attitudes [60]. Technology competency and the type of information to share are also connected to privacy concerns [37,89]. Nevertheless, what cannot be neglected is that technology itself cannot handle upcoming challenges. A more holistic approach should also include proactive legal and ethical guidelines in order to increase user data sharing willingness as well as secure data exchange between health care systems in the future.

Limitations

The interpretation of data sharing attitudes is a complex process related to and influenced by various factors [84] including the historical or regulatory context of data sharing in each nation [81]. In this study, we aimed to explore and present an overview of the data sharing perspectives adopting a quantitative approach. The research approach of this paper is related to the scope of the IHAN project and aims to present an overview of user data sharing attitudes in four European countries; therefore, the generalization of results is limited in this scope. We suggest that future qualitative or mixed methods approaches could articulate, distill, and contextualize knowledge in relation to the causality of user willingness toward sharing PHD. The analysis and discussion of the results in this study relied only on descriptive statistics and visualizations. Combined with the quantitative nature of the research, it minimizes threats to the validity of the conclusions. The inclusion of a representative sample in relation to gender, age, and locality also mitigates the risk of sampling bias. In addition, the bias related to sampling is minimum due to the random selection of the study participants. The high response rates can probably be attributed to the financial compensations that participants have received [90]. A methodological limitation, which is common in Web-based surveys, concerns the participation rate and atypical timestamp [91]. The lack of information about these two metrics raises implications about whether the participants are representative of the population intended to be analyzed and about the quality of responses in relation to the time elapsed between the dispatch of the survey and its completion. Another consideration regards a limitation that can arise from the translation and back-translation of the questionnaire and results. Although no pilot testing has been performed to validate the translated items, the development of the questionnaire by specialists, use of a consistent team of professional translators throughout the project, and quantitative nature of the study has minimized the risk.

Conclusions

Our findings reinforce the idea that the majority of connected health users are willing to share their PHD and prefer to do so for scientific research. Furthermore, this study highlights that age, education, and occupation as well as the level of digitalization in the country are significant factors affecting user data sharing attitudes. Positive attitudes toward scientific research highlighted by our results can be perceived as an indication of trusted relations between users and academia.

Our results have both theoretical and practical implications. The enforcement of the new GDPR in Europe demands more

active participation of users in the management of their personal data. Nevertheless, this requires users to be knowledgeable and fully comprehend their rights and options over privacy settings. Educating users on the benefits and safety of data sharing could facilitate fair handling of data and increase trust. From the perspective of service providers, future efforts need to be directed toward simplification of privacy statements and reconsideration of systems design. The provision of user-friendly interfaces to enable faster and seamless screening of privacy settings could be another practical implication for technology developers. From the standpoint of policy makers, this study can facilitate shifts in policies toward accommodating user needs more effectively. Educating users, information technology professionals, and service providers on the importance of implementing public policy that supports responsibility of user-mediators while identifying gaps in policy frameworks are some of the steps that stakeholders should consider in the creation of future health care systems.

Acknowledgments

This article is based upon work from COST Action ENJECT TD 1405, supported by COST (European Cooperation in Science and Technology; www.cost.eu). We would like to thank Sitra Innovation Fund for providing us with the data.

Authors' Contributions

All authors conceived the study and led interpretation of the data. MK authored and revised the manuscript. SO carried out the curation of visualizations. SO and MI suggested critical modifications in the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Completed Checklist for Reporting Results of Internet E-Surveys.
[DOCX File , 14 KB - jmir_v21i11e14537_app1.docx]

Multimedia Appendix 2
Questionnaire survey results.
[DOCX File , 23 KB - jmir_v21i11e14537_app2.docx]

References

1. Moghaddasi H, Asadi F, Hosseini A, Ebnehoseini Z. E-Health: a global approach with extensive semantic variation. J Med Syst 2012 Oct;36(5):3173-3176. [doi: 10.1007/s10916-011-9805-z] [Medline: 22113437]
2. Riazul Islam SM, Kwak D, Kabir M, Hossain M, Kwak KS. The Internet of Things for health care: a comprehensive survey. IEEE Access 2015;3:678-708. [doi: 10.1109/ACCESS.2015.2437951]
3. Willcox JC, van der Pligt P, Ball K, Wilkinson SA, Lappas M, McCarthy EA, et al. Views of women and health professionals on mHealth lifestyle interventions in pregnancy: a qualitative investigation. JMIR Mhealth Uhealth 2015 Oct 28;3(4):e99 [FREE Full text] [doi: 10.2196/mhealth.4869] [Medline: 26510886]
4. International Data Corporation. 2019 Jun 18. Smartphone market share URL: https://www.idc.com/promo/smartphone-market-share/os [accessed 2019-09-01]
5. Zargaran A, Ash J, Kerry G, Rasasingam D, Gokani S, Mittal A, et al. Ethics of smartphone usage for medical image sharing. Indian J Surg 2018 Jun 30;80(3):300-301 [FREE Full text] [doi: 10.1007/s12262-017-1710-3] [Medline: 29973771]
6. Evans D. Cisco Internet Business Solutions Group. 2011. The Internet of Things: how the next evolution of the internet is changing everything URL: https://www.cisco.com/c/dam/en_us/about/ac79/docs/innov/IoT_IBSG_0411FINAL.pdf [accessed 2019-09-01]
7. Wang Y, Kung L, Wang WYC, Cegielski CG. An integrated big data analytics-enabled transformation model: application to health care. Inf Manag 2018 Jan;55(1):64-79. [doi: 10.1016/j.im.2017.04.001]
8. Karampela M, Grundstrom C, Isomursu M. Personal health data: access and perceived value in Denmark. 2018 Presented at: Proceedings of the 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society; 2018; Honolulu p. 4081-4084. [doi: 10.1109/embc.2018.8513407]
9. Vergados DD. Service personalization for assistive living in a mobile ambient healthcare-networked environment. Pers Ubiquit Comput 2010 Mar 3;14(6):575-590. [doi: 10.1007/s00779-009-0278-8]
10. Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. Health Inf Sci Syst 2014;2:3 [FREE Full text] [doi: 10.1186/2047-2501-2-3] [Medline: 25825667]

11. Taylor K. Connected health: how digital technology is transforming health and social care. London: Deloitte Centre for Healthcare Solutions; 2010. URL: https://www2.deloitte.com/content/dam/Deloitte/uk/Documents/life-sciences-health-care/deloitte-uk-connected-health.pdf [accessed 2019-10-21]
12. Carroll N. Key success factors for smart and connected health software solutions. Computer 2016 Nov;49(11):22-28. [doi: 10.1109/mc.2016.340]
13. Caulfield BM, Donnelly SC. What is connected health and why will it change your practice? QJM 2013 Aug;106(8):703-707. [doi: 10.1093/qjmed/hct114] [Medline: 23676416]
14. Steinhubl S, Muse E, Topol E. Can mobile health technologies transform health care? JAMA 2013 Dec 11;310(22):2395-2396. [doi: 10.1001/jama.2013.281078] [Medline: 24158428]
15. DuBenske LL, Gustafson DH, Shaw BR, Cleary JF. Web-based cancer communication and decision making systems: connecting patients, caregivers, and clinicians for improved health outcomes. Med Decis Making 2010;30(6):732-744 [FREE Full text] [doi: 10.1177/0272989X10386382] [Medline: 21041539]
16. Swan M. Emerging patient-driven health care models: an examination of health social networks, consumer personalized medicine and quantified self-tracking. Int J Environ Res Public Health 2009 Feb;6(2):492-525 [FREE Full text] [doi: 10.3390/ijerph6020492] [Medline: 19440396]
17. Ahern DK, Kreslake JM, Phalen JM. What is eHealth (6): perspectives on the evolution of eHealth research. J Med Internet Res 2006 Mar 31;8(1):e4 [FREE Full text] [doi: 10.2196/jmir.8.1.e4] [Medline: 16585029]
18. Mirza F, Norris T, Stockdale R. Mobile technologies and the holistic management of chronic diseases. Health Informatics J 2008 Dec;14(4):309-321. [doi: 10.1177/1460458208096559] [Medline: 19008280]
19. Karampela M, Ouhbi S, Isomursu M. Personal health data: a systematic mapping study. Int J Med Inform 2018 Oct;118:86-98. [doi: 10.1016/j.ijmedinf.2018.08.006] [Medline: 30153927]
20. Karampela M, Ouhbi S, Isomursu M. Accessing and sharing electronic personal health data. 2019 Presented at: Proceedings of the 12th International Joint Conference on Biomedical Engineering Systems and Technologies; 2019; Prague p. 182-189.
21. Ouhbi S, Fernández-Alemán J, Toval A, Rivera Pozo J, Idri A. Sustainability requirements for connected health applications. J Softw Evol Proc 2017 Nov 28;30(7):e1922. [doi: 10.1002/smr.1922]
22. Flu Near You. URL: https://flunearyou.org/#!/ [accessed 2019-04-01]
23. Healthmap. URL: https://www.healthmap.org/en/ [accessed 2019-04-01]
24. Wicks P, Massagli M, Frost J, Brownstein C, Okun S, Vaughan T, et al. Sharing health data for better outcomes on PatientsLikeMe. J Med Internet Res 2010 Jun 14;12(2):e19 [FREE Full text] [doi: 10.2196/jmir.1549] [Medline: 20542858]
25. Wicks P, Keininger DL, Massagli MP, de la Loge C, Brownstein C, Isojärvi J, et al. Perceived benefits of sharing health data between people with epilepsy on an online platform. Epilepsy Behav 2012 Jan;23(1):16-23 [FREE Full text] [doi: 10.1016/j.yebeh.2011.09.026] [Medline: 22099528]
26. Green AK, Reeder-Hayes KE, Corty RW, Basch E, Milowsky MI, Dusetzina SB, et al. The project data sphere initiative: accelerating cancer research by sharing data. Oncologist 2015 May;20(5):464 [FREE Full text] [doi: 10.1634/theoncologist.2014-0431] [Medline: 25876994]
27. Lo B. Sharing clinical trial data: maximizing benefits, minimizing risk. JAMA 2015 Feb 24;313(8):793-794. [doi: 10.1001/jama.2015.292] [Medline: 25594500]
28. Sousa VD, Ramalho PI, Silveira D. Sharing regulatory data as tools for strengthening health systems in the Region of the Americas. Rev Panam Salud Publica 2016 May;39(5):245-254. [Medline: 27706398]
29. European Medicines Agency. Publication of clinical data. URL: https://www.ema.europa.eu/en/human-regulatory/marketing-authorisation/clinical-data-publication [accessed 2019-04-01]
30. Aboelfotoh M, Martin P, Hassanein H. A mobile-based architecture for integrating personal health record data. 2014 Presented at: 16th International Conference on e-Health Networking, Applications and Services (Healthcom); 2014; Natal p. 269-274. [doi: 10.1109/HealthCom.2014.7001853]
31. Vahidhunnisha J, Balasubramaniam T, Ramas A. Improving privacy in sharing of personal health data storage on cloud. Int J Innov Technol Res 2014;2(3):944.
32. Fishbein M, Ajzen I. Belief, Attitude, Intention and Behavior: An Introduction to Theory and Research. Boston: Addison-Wesley Pub Co; 1975.
33. Karampela M, Ouhbi S, Isomurso M. Exploring users' willingness to share their health and personal data under the prism of the new GDPR: implications in healthcare. 2019 Presented at: 41st Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC); 2019; Berlin.
34. Ajzen I. The theory of planned behavior. Organ Behav Hum Decis Process 1991 Dec;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-T]
35. Davis F, Warshaw P. What do intention scales measure? J Gen Psychol 1992 Oct;119(4):391-407. [doi: 10.1080/00221309.1992.9921181]
36. Athey S, Catalini C, Tucker C. The digital privacy paradox: small money, small costs, small talk. SSRN J 2018 Apr 08:1. [doi: 10.2139/ssrn.2916489]
37. Pickard KT, Swan M. Big desire to share big health data: a shift in consumer attitudes toward personal health information. 2014. URL: https://pdfs.semanticscholar.org/b8ae/3f8efdec637bce3ea3844695e59d215421d2.pdf [accessed 2019-10-21]

38. Chen J, Bauman A, Allman-Farinelli M. A study to determine the most popular lifestyle smartphone applications and willingness of the public to share their personal data for health research. Telemed J E Health 2016 Aug;22(8):655-665. [doi: 10.1089/tmj.2015.0159] [Medline: 26958742]
39. 23andMe: DNA genetic testing & analysis. URL: https://www.23andme.com/en-eu/ [accessed 2019-04-01]
40. PatientsLikeMe. URL: https://www.patientslikeme.com/ [accessed 2019-04-01]
41. Whiddett R, Hunter I, Engelbrecht J, Handy J. Patients' attitudes towards sharing their health information. Int J Med Inform 2006 Jul;75(7):530-541. [doi: 10.1016/j.ijmedinf.2005.08.009] [Medline: 16198142]
42. Agarwal R, Anderson C. The complexity of consumer willingness to disclose personal information: unraveling health information privacy concerns. 2008 Presented at: eHealth Initiative 5th Annual Conference; 2008; Portoroz.
43. Hartzler A, Skeels MM, Mukai M, Powell C, Klasnja P, Pratt W. Sharing is caring, but not error free: transparency of granular controls for sharing personal health information in social networks. AMIA Annu Symp Proc 2011;2011:559-568 [FREE Full text] [Medline: 22195111]
44. Weitzman ER, Kelemen S, Kaci L, Mandl KD. Willingness to share personal health record data for care improvement and public health: a survey of experienced personal health record users. BMC Med Inform Decis Mak 2012 May 22;12:39 [FREE Full text] [doi: 10.1186/1472-6947-12-39] [Medline: 22616619]
45. Medford-Davis LN, Chang L, Rhodes KV. Health information exchange: what do patients want? Health Informatics J 2017 Dec;23(4):268-278. [doi: 10.1177/1460458216647190] [Medline: 27245671]
46. Weng C, Friedman C, Rommel C, Hurdle J. A two-site survey of medical center personnel's willingness to share clinical data for research: implications for reproducible health NLP research. BMC Med Inform Decis Mak 2019 Apr 04;19(Suppl 3):70 [FREE Full text] [doi: 10.1186/s12911-019-0778-z] [Medline: 30943963]
47. Kaye J, Curren L, Anderson N, Edwards K, Fullerton SM, Kanellopoulou N, et al. From patients to partners: participant-centric initiatives in biomedical research. Nat Rev Genet 2012 May;13(5):371-376 [FREE Full text] [doi: 10.1038/nrg3218] [Medline: 22473380]
48. Kim K, Sankar P, Wilson M, Haynes S. Factors affecting willingness to share electronic health data among California consumers. BMC Med Ethics 2017;18(1):25. [Medline: 28376801]
49. Weitzman ER, Kaci L, Mandl KD. Sharing medical data for health research: the early personal health record experience. J Med Internet Res 2010 May 25;12(2):e14 [FREE Full text] [doi: 10.2196/jmir.1356] [Medline: 20501431]
50. Audrey S, Brown L, Campbell R, Boyd A, Macleod J. Young people's views about consenting to data linkage: findings from the PEARL qualitative study. BMC Med Res Methodol 2016;16(1):34. [Medline: 27001504]
51. Buckley BS, Murphy AW, MacFarlane AE. Public attitudes to the use in research of personal health information from general practitioners' records: a survey of the Irish general public. J Med Ethics 2011 Jan;37(1):50-55. [doi: 10.1136/jme.2010.037903] [Medline: 21071570]
52. Haddow G, Bruce A, Sathanandam S, Wyatt JC. 'Nothing is really safe': a focus group study on the processes of anonymizing and sharing of health data for research purposes. J Eval Clin Pract 2011 Dec;17(6):1140-1146. [doi: 10.1111/j.1365-2753.2010.01488.x] [Medline: 20629997]
53. Papoutsi C, Reed JE, Marston C, Lewis R, Majeed A, Bell D. Patient and public views about the security and privacy of Electronic Health Records (EHRs) in the UK: results from a mixed methods study. BMC Med Inform Decis Mak 2015 Oct 14;15:86 [FREE Full text] [doi: 10.1186/s12911-015-0202-2] [Medline: 26466787]
54. Anderson C, Agarwal R. The digitization of healthcare: boundary risks, emotion, and consumer willingness to disclose personal health information. Info Sys Res 2011 Sep;22(3):469-490. [doi: 10.1287/isre.1100.0335]
55. Dinev T, Hart P. An extended privacy calculus model for e-commerce transactions. Info Sys Res 2006 Mar;17(1):61-80. [doi: 10.1287/isre.1060.0080]
56. Shah N, Coathup V, Teare H, Forgie I, Giordano GN, Hansen TH, et al. Motivations for data sharing-views of research participants from four European countries: a DIRECT study. Eur J Hum Genet 2019 May;27(5):721-729. [doi: 10.1038/s41431-019-0344-2] [Medline: 30700834]
57. Stockdale J, Cassell J, Ford E. “Giving something back”: a systematic review and ethical enquiry into public views on the use of patient data for research in the United Kingdom and the Republic of Ireland. Wellcome Open Res 2018;3:6 [FREE Full text] [doi: 10.12688/wellcomeopenres.13531.2] [Medline: 30854470]
58. Becker M. Understanding users' health information privacy concerns for health wearables. URL: https://core.ac.uk/download/pdf/143481249.pdf [accessed 2019-10-21]
59. Abdelhamid M, Gaia J, Sanders GL. Putting the focus back on the patient: how privacy concerns affect personal health information sharing intentions. J Med Internet Res 2017 Sep 13;19(9):e169 [FREE Full text] [doi: 10.2196/jmir.6877] [Medline: 28903895]
60. Angst C, Agarwal R. MISQ. 2009. Adoption of electronic health records in the presence of privacy concerns: the elaboration likelihood model and individual persuasion URL: https://www3.nd.edu/~cangst/CoreyAngst_FacultyWebsite_files/Angst2009MISQ.pdf [accessed 2019-10-21]
61. Harjumaa M, Saraniemi S, Pekkarinen S, Lappi M, Similä H, Isomursu M. Feasibility of digital footprint data for health analytics and services: an explorative pilot study. BMC Med Inform Decis Mak 2016 Nov 09;16(1):139. [doi: 10.1186/s12911-016-0378-0] [Medline: 27829413]

62. Girardin F, Calabrese F, Fiore F, Ratti C, Blat J. Digital Footprinting: Uncovering Tourists With User-Generated Content. Piscataway: Institute of Electrical and Electronics Engineers; 2008.
63. IHAN proof of concept pilot. URL: https://www.sitra.fi/en/projects/ihan-proof-concept-pilots/ [accessed 2019-04-01]
64. McCormack P, Kole A, Gainotti S, Mascalzoni D, Molster C, Lochmüller H, et al. 'You should at least ask'. The expectations, hopes and fears of rare disease patients on large-scale data and biomaterial sharing for genomics research. Eur J Hum Genet 2016 Oct;24(10):1403-1408 [FREE Full text] [doi: 10.1038/ejhg.2016.30] [Medline: 27049302]
65. Courbier S, Dimond R, Bros-Facer V. Share and protect our health data: an evidence based approach to rare disease patients' perspectives on data sharing and data protection: quantitative survey and recommendations. Orphanet J Rare Dis 2019 Jul 12;14(1):175 [FREE Full text] [doi: 10.1186/s13023-019-1123-4] [Medline: 31300010]
66. Special Eurobarometer 359: attitudes on data protection and electronic identity in the European Union. URL: http://data.europa.eu/euodp/data/dataset/S864_74_3_EBS359 [accessed 2019-09-01]
67. Gaupp-Berghausen M, Raser E, Anaya-Boig E, Avila-Palencia I, de Nazelle A, Dons E, et al. Evaluation of different recruitment methods: longitudinal, web-based, pan-European physical activity through sustainable transport approaches (PASTA) project. J Med Internet Res 2019 Mar 09;21(5):e11492 [FREE Full text] [doi: 10.2196/11492] [Medline: 31066715]
68. Eysenbach G. Improving the quality of Web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
69. Shah N, Coathup V, Teare H, Forgie I, Giordano GN, Hansen TH, et al. Sharing data for future research-engaging participants' views about data governance beyond the original project: a DIRECT Study. Genet Med 2019 May;21(5):1131-1138. [doi: 10.1038/s41436-018-0299-7] [Medline: 30262927]
70. Kowalewski S, Ziefle M, Ziegeldorf H, Wehrle K. Like us on Facebook! Analyzing user preferences regarding privacy settings in Germany. Procedia Manufactur 2015;3:815-822. [doi: 10.1016/j.promfg.2015.07.336]
71. Choi NG, Dinitto DM. The digital divide among low-income homebound older adults: internet use patterns, eHealth literacy, and attitudes toward computer/Internet use. J Med Internet Res 2013 May;15(5):e93 [FREE Full text] [doi: 10.2196/jmir.2645] [Medline: 23639979]
72. Ziefle M, Halbey J, Kowalewski S. Users willingness to share data on the internet: perceived benefits and caveats. 2016 Presented at: Proceedings of the International Conference on Internet of Things and Big Data; 2016; Melbourne p. 255-265. [doi: 10.5220/0005897402550265]
73. Akhter HS. Privacy concern and online transactions: the impact of internet self-efficacy and internet involvement. J Consum Mark 2014 May 06;31(2):118-125. [doi: 10.1108/jcm-06-2013-0606]
74. Fogel J, Nehmad E. Internet social network communities: risk taking, trust, and privacy concerns. Comput Hum Behav 2009 Jan;25(1):153-160. [doi: 10.1016/j.chb.2008.08.006]
75. Freestone O, Mitchell V. Generation Y attitudes towards E-ethics and internet-related misbehaviours. J Buss Ethics 2004 Oct;54(2):121-128. [doi: 10.1007/s10551-004-1571-0]
76. Wilson M, Scalise K, Gochyyev P. Rethinking ICT literacy: from computer skills to social network settings. Think Skills Creativ 2015 Dec;18:65-80. [doi: 10.1016/j.tsc.2015.05.001]
77. Digital economy and society index report 2018. URL: http://ec.europa.eu/newsroom/dae/document.cfm?doc_id=52244 [accessed 2019-04-01]
78. Global open data index. 2019. URL: https://index.okfn.org [accessed 2019-04-01]
79. Schwartz P. European data protection law and restrictions on international data flows. Iowa L Rev 1994;80:471.
80. Righettini MS. Institutionalization, leadership, and regulative policy style: a France/Italy comparison of data protection authorities. J Compar Pol Analys 2011 Apr;13(2):143-164. [doi: 10.1080/13876988.2011.555995]
81. Smit EG, Van Noort G, Voorveld HA. Understanding online behavioural advertising: user knowledge, privacy concerns and online coping behaviour in Europe. Comput Hum Behav 2014 Mar;32:15-22. [doi: 10.1016/j.chb.2013.11.008]
82. Jalali MS, Kaiser JP. Cybersecurity in hospitals: a systematic, organizational perspective. J Med Internet Res 2018 Dec 28;20(5):e10059 [FREE Full text] [doi: 10.2196/10059] [Medline: 29807882]
83. Pickard K. Exploring markets of data for personal health information. 2014 Presented at: IEEE International Conference on Data Mining Workshop; 2014; Shenzhen p. 477-480. [doi: 10.1109/ICDMW.2014.120]
84. Fox S, Purcell K. Chronic disease and the internet. Washington: Pew Internet & American Life Project; 2010. URL: https://www.pewinternet.org/2010/03/24/chronic-disease-and-the-internet/ [accessed 2019-09-01]
85. Lewis K, Kaufman JN. The taste for privacy: an analysis of college student privacy settings in an online social network. J Computer-Mediated Commun 2008;14(1):79-100. [doi: 10.1111/j.1083-6101.2008.01432.x]
86. Utz S, Kramer N. The privacy paradox on social network sites revisited: the role of individual characteristics and group norms. J Psychosoc Res Cyberspace 2009;3(2):1.
87. Gymrek M, McGuire AL, Golan D, Halperin E, Erlich Y. Identifying personal genomes by surname inference. Science 2013 Jan 18;339(6117):321-324 [FREE Full text] [doi: 10.1126/science.1229566] [Medline: 23329047]
88. Richter G, Borzikowsky C, Lieb W, Schreiber S, Krawczak M, Buyx A. Patient views on research use of clinical data without consent: legal, but also acceptable? Eur J Hum Genet 2019 Jun;27(6):841-847. [doi: 10.1038/s41431-019-0340-6] [Medline: 30683927]
89. Han P, Maclaurin A. Do consumers really care about online privacy? Mark Manag 2001;11(1):35.

90. Parkinson A, Jorm L, Douglas KA, Gee A, Sargent GM, Lujic S, et al. Recruiting general practitioners for surveys: reflections on the difficulties and some lessons learned. Aust J Prim Health 2015;21(2):254-258. [doi: 10.1071/PY13129] [Medline: 24581281]
91. Karlsen MC, Lichtenstein AH, Economos CD, Folta SC, Rogers G, Jacques PF, et al. Web-based recruitment and survey methodology to maximize response rates from followers of popular diets: the adhering to dietary approaches for personal taste (ADAPT) feasibility survey. Curr Dev Nutr 2018 May;2(5):nzy012 [FREE Full text] [doi: 10.1093/cdn/nzy012] [Medline: 29955724]

Abbreviations

CHERRIES: Checklist for Reporting Results of Internet E-Surveys
EU: European Union
GDPR: General Data Protection Regulation
IoT: Internet of Things
PHD: personal health data
RQ: research question

Edited by M Tilney; submitted 30.04.19; peer-reviewed by C Jacob; comments to author 10.07.19; revised version received 19.07.19; accepted 09.10.19; published 27.11.19.

Please cite as:
Karampela M, Ouhbi S, Isomursu M
Connected Health User Willingness to Share Personal Health Data: Questionnaire Study
J Med Internet Res 2019;21(11):e14537
URL: https://www.jmir.org/2019/11/e14537
doi:10.2196/14537
PMID:31774410

Original Paper

Electronic Screening for Alcohol Use and Brief Intervention by Email for University Students: Reanalysis of Findings From a Randomized Controlled Trial Using a Bayesian Framework

Marcus Bendtsen[1], PhD

Department of Medical and Health Sciences, Linköping University, Linköping, Sweden

Corresponding Author:
Marcus Bendtsen, PhD
Department of Medical and Health Sciences
Linköping University
Linköping, 58183
Sweden
Phone: 46 13281000
Email: marcus.bendtsen@liu.se

Abstract

Background: Almost a decade ago, Sweden became the first country to implement a national system enabling student health care centers across all universities to routinely administer (via email) an electronic alcohol screening and brief intervention to their students. The Alcohol email assessment and feedback study dismantling effectiveness for university students (AMADEUS-1) trial aimed to assess the effect of the student health care centers' routine practices by exploiting the lack of any standard timing for the email invitation and by masking trial participation from students. The original analyses adopted the conventional null hypothesis framework, and the results were consistently in the expected direction. However, since for some tests the *P* values did not pass the conventional .05 threshold, some of the analyses were necessarily inconclusive.

Objective: The outcomes of the AMADEUS-1 trial were derived from the first 3 items of the Alcohol Use Disorders Identification Test (AUDIT-C). The aim of this paper was to reanalyze the two primary outcomes of the AMADEUS-1 trial (AUDIT-C scores and prevalence of risky drinking), using the same models used in the original publication but applying a Bayesian inference framework and interpretation.

Methods: The same regression models used in the original analysis were employed in this reanalysis (linear and logistic regression). Model parameters were given uniform priors. Markov chain Monte Carlo was used for Bayesian inference, and posterior probabilities were calculated for prespecified thresholds of interest.

Results: Where the null hypothesis tests showed inconclusive results, the Bayesian analysis showed that offering an intervention at baseline was preferable compared to offering nothing. At follow-up, the probability of a lower AUDIT-C score among those who had been offered an intervention at baseline was greater than 95%, as was the case when comparing the prevalence of risky drinking.

Conclusions: The Bayesian analysis allows for a more consistent perspective of the data collected in the trial, since dichotomization of evidence is not looked for at some arbitrary threshold. Results are presented that represent the data collected in the trial rather than trying to make conclusions about the existence of a population effect. Thus, policy makers can think about the value of keeping the national system without having to navigate the treacherous landscape of statistical significance.

Trial Registration: ISRCTN Registry ISRCTN28328154; http://www.isrctn.com/ISRCTN28328154

(J Med Internet Res 2019;21(11):e14419) doi:10.2196/14419

KEYWORDS

Bayesian analysis; telemedicine; alcohol; randomized controlled trial

Introduction

Background

Alcohol consumption contributed to more than 4.5% of deaths globally in 2016 [1] and was the leading risk factor among the population aged 15-49 years old. It has been suggested that alcohol policies might need to be revised worldwide to lower the overall population-level consumption [2]. While policies controlling price and availability may be one way forward [3], the advent of electronic health (eHealth) interventions has made us better equipped to deliver personal behavior change interventions to larger populations.

Early initiatives to use digital means of delivering alcohol interventions came in the form of electronic screening and brief interventions (eSBIs) [4-8]. Typically, these interventions ask participants to complete a questionnaire, after which feedback is given on their responses and some advice on behavior change is offered (based on recommended drinking levels). The feedback and advice are commonly designed around behavior change theories and models, such as protection motivation theory [9], social cognitive theory [10] and the theory of planned behavior [11]. In general, eHealth interventions for alcohol behavior change have shown promise when they have included components that focus on behavior substitution, problem solving, goal setting, review of behavioral goals, self-monitoring, and normative feedback [12,13].

Meta-analyses suggest that there exists a small positive effect of eSBIs on the amount of alcohol consumed weekly in the short term, with a Cohen d=−0.17 (95% CI −0.27 to −0.18) found in one analysis [14], a Cohen d=−0.14 (95% CI −0.24 to −0.03) in another analysis [15], and a weighted mean difference of alcohol in grams=−16.59 (95% CI −23.70 to −9.48) in a third analysis [16]. Although long-term effects have not been measurable, these brief interventions are nevertheless useful for reaching many individuals at a low cost.

Almost a decade ago, Sweden became the first country to implement a national system enabling student health care centers across all universities to routinely administer eSBIs. The system, which is still routinely used today, sends an email to all university students with an invitation and a hyperlink to a 10-item web questionnaire which is then followed by personal feedback and advice. At the time this system was introduced, there was some evidence of the effectiveness of eSBIs but there was a paucity for large-scale, multisite, effectiveness trials of routine care systems.

The Alcohol Email Assessment and Feedback Study Dismantling Effectiveness for University Students Trial

The Alcohol email assessment and feedback study dismantling effectiveness for university students (AMADEUS-1) trial [4,7], conducted in 2011, aimed to assess the effect of the student health care centers' routine practices by exploiting the lack of any standard timing of the email invitation and by masking trial participation from students. The trial outcome was originally reported in 2013 [7].

During the autumn term of 2011, all students in semesters 1, 3 and 5 at two universities in Sweden (Linköping and Luleå) were included in the AMADEUS-1 trial. Notably, students' email addresses were randomized into 3 groups (Group 1, Group 2, and Group 3) prior to any invitation or contact with the students. Ethical concerns with the use of this type of masking was considered and approved by the Regional Ethical Committee in Linköping, Sweden (No 2010/291-31). In a subsequent trial (the AMADEUS-2 trial [6,8]) a more conventional approach was used to estimate the effect of eSBIs on harmful and hazardous drinkers. A Bayesian reanalysis of the AMADEUS-2 trial has also been reported [17].

On September 5, 2011, Group 1 and Group 2 were sent an email from the student health care center with a hyperlink to a web questionnaire comprising 10 items which assessed their current alcohol consumption, masked as part of routine care. Group 1 was additionally told that they would also get feedback, which they received immediately after responding to the questionnaire. Group 2 was thanked for their participation and offered a hyperlink to a website with general information about alcohol, which was not believed to have any content helpful for supporting behavior change. Group 3 was not contacted at this time.

Three months after the initial email to Group 1 and Group 2, all three groups were sent identical emails with an invitation to participate in a web-based general lifestyle survey where 3 out of the 15 items were the first 3 items of the Alcohol Use Disorders Identification Test (AUDIT-C [18]). Crucially, this invitation made no reference to the alcohol assessment conducted three months earlier and it was not disclosed as a follow-up questionnaire in a randomized trial.

Objectives

Outcomes of the AMADEUS-1 trial were derived from the 3 AUDIT-C items in the general lifestyle survey. This reanalysis will focus on two primary outcomes: AUDIT-C scores and prevalence of risky drinking. In Sweden, risky drinkers are those who fulfil at least one of two criteria: (1) heavy episodic drinking of at least 4 (female) or 5 (male) standard drinks of alcohol on one occasion the past month; or (2) consuming more than 9 (female) or 14 (male) standard drinks of alcohol per week. One standard drink is defined as 12 grams of alcohol in Sweden.

The current goal is to reanalyze the two primary outcomes of the AMADEUS-1 trial, using the same models used in the original publication but also using a Bayesian inference framework and interpretation.

Methods

Overview

In the original analysis of the AMADEUS-1 trial, normal regression was used to contrast AUDIT-C scores (log-transformed) and logistic regression was used to contrast risky drinking. Both models were adjusted for baseline variables. In this Bayesian analysis, the same regression models were used, and uniform priors were applied to all model parameters. The full specifications of the Bayesian models can be seen in

the following 2 equations. Separate analyses were done comparing Group 1 versus Group 3 and Group 2 versus Group 3. In all cases, Group 3 was considered the control group and Group 1 and Group 2 were considered intervention groups.

Equation 1:

Equation 2:

Alcohol Use Disorders Identification Test

When contrasting AUDIT-C scores (Equation 1), the primary interest of the analyses was the regression coefficient for the group variable (α_1). A negative value for α_1 suggests that the group which was randomized to receive an intervention (Group 1 and Group 2 respectively) had, on average, lower AUDIT-C scores at follow-up than the group which was randomized to the control setting (Group 3). Coefficients were back transformed prior to inspection. Informed by the original analysis, it was decided that thresholds of interest for which the marginal posterior distribution for α_1 should be inspected were 0, –0.02, and –0.04. The thresholds were chosen to communicate whether offering an intervention is preferable to not doing so (the 0 threshold), and to indicate the magnitude of the difference between groups (–0.02 and –0.04).

Risky Drinking

When contrasting risky drinking (Equation 2), the primary interest was the regression coefficient for the group variable (β_1), that is the log of the odds ratio (OR) between the group which was randomized to an intervention (Group 1 and Group 2 respectively) and the group which was randomized to the control setting (Group 3). Coefficients were exponentiated before inspection, thus a value of β_1 lower than 1 would suggest that the odds of risky drinking in the intervention group was lower than the odds in the control. Informed by the original analysis, it was decided that thresholds of interest for which the marginal posterior distribution for β_1 should be inspected was 1, 0.9 and 0.8. Again, the thresholds were chosen to communicate whether offering an intervention is preferable to not doing so (the 1 threshold), and to indicate the magnitude of the difference between the groups (0.9 and 0.8).

Inference

Markov chain Monte Carlo was used for Bayesian inference (RStan version 2.16.2). For each model, 50,000 iterations were run with 25,000 warmup iterations in four chains. Inference for AUDIT-C scores (Equation 1) took 3.5 minutes, and for risky drinking (Equation 2) 5.5 minutes. All computations were done on a MacBook Pro (2017 model).

Results

Primary Findings

A total of 14,910 students were randomized into the 3 arms of the trial. In Group 1, 36.2% (1798/4969) of participants completed the eSBI, 32.6% (1621/4969) of Group 2 participants completed the alcohol screening questionnaire, and as previously discussed Group 3 was not contacted at this point. Approximately half of all students responded to the general lifestyle survey that was sent three months after randomization: 51.2% (2546/4969) in Group 1, 52.2% (2594/4969) in Group 2, and 53.7% (2669/4972) in Group 3.

Original Analysis: Null Hypothesis Framework

The original analysis for the AMADEUS-1 trial is presented in Table 1 [7]. Null hypothesis tests were two-tailed and assessed at the .05 threshold. It was found that Group 1 and Group 3 did not report a statistically significant difference with respect to AUDIT-C scores (*P*=.07) while Group 2 and Group 3 did (*P*=.04), with Group 2, on average, reporting a lower AUDIT-C score than Group 3. Risky drinking was found to be statistically significantly different between Group 1 and Group 3 (*P*=.006), with risky drinking less prevalent in Group 1 than in Group 3, but not so for Group 2 and Group 3 (*P*=.08).

As a reminder, *P* values indicate how likely it is that we would have seen the data that we did in the trial in a hypothetical world where the population effect is exactly zero. Convention says that if the data is less likely than 5%, then we should reject the hypothetical world. However, it does not mean that if the 5% threshold is not broken that we should accept the hypothetical world. Instead, the confidence interval indicates which hypothetical worlds cannot be rejected given the data that we have seen in the trial.

Table 1. Original analysis of AUDIT-C and risky drinking at follow-up, comparing Group 1 versus 2 and Group 2 versus 3.

Categories	Group 1 (n=2546)	Group 2 (n=2594)	Group 3 (n=2669)	Group 1 versus 3		Group 2 versus 3	
				Regression coefficient[a], 95% CI	*P* value	Regression coefficient[a], 95% CI	*P* value
AUDIT-C[b]	3.46 (3.09)[c]	3.44 (3.17)[c]	3.60 (3.14)[c]	–0.032 (–0.066 to 0.003)	.07	–0.038 (–0.072 to –0.002)	.04
Risky drinking[d]	1136 (44.6)[e]	1194 (46.0)[e]	1288 (48.3)[e]	0.85 (0.76 to 0.95)	.006	0.90 (0.81 to 1.01)	.08

[a]Linear coefficient for AUDIT-C scores (back transformed) and odds ratio for risky drinking (adjusted for sex, age, university, and semester).

[b]AUDIT-C: Alcohol Use Disorders Identification Test.

[c]Geometric mean (SD). Approximate standard deviation back-calculated from the log-scale.

[d]Risky drinking: heavy episodic drinking ≥1 a month or weekly consumption >14 for men and >9 for women (Swedish national guidelines).

[e]n (%).

Bayesian Analysis

The computational result of a Bayesian analysis using Markov chain Monte Carlo uses samples from the posterior distribution of each parameter of interest. Histograms of these samples are shown in Figures 1-4 for the coefficient for the group variable in the AUDIT-C models (α_1 in Equation 1, back transformed) and the risky drinking models (β_1 in Equation 2, exponentiated). For instance, in Figure 1 we can see that there is a majority of samples to the left of 0, indicating that it is more likely than not that there was a difference in AUDIT-C scores at follow-up between Groups 1 and 3. Similarly, in Figure 4 we can see that a majority of the samples are to the left of 1, indicating that the prevalence of risky drinking in Group 2 was lower than in Group 3 at follow-up (ie, the OR was lower than 1). For the enclosed analyses, no trends were found in the sampling when inspecting trace plots (see Multimedia Appendix 1).

Rather than just visually inspecting the histograms, the samples drawn during inference can be used to calculate probabilities of interest (Tables 2 and 3). For example, when comparing the prevalence of risky drinking between Group 1 and Group 3 in Table 3, the ratio of samples that were lower than 1 was 99.7%, thus there was a 99.7% probability that the OR was less than 1 (indicating fewer risky drinkers in Group 1 compared to Group 3). Furthermore, there was an 82.4% probability that the OR was less than 0.9.

Figure 1. Samples from the posterior distribution of α_1 in the AUDIT-C model when comparing Group 1 versus Group 3 (Equation 1, back transformed). AUDIT-C: Alcohol Use Disorders Identification Test.

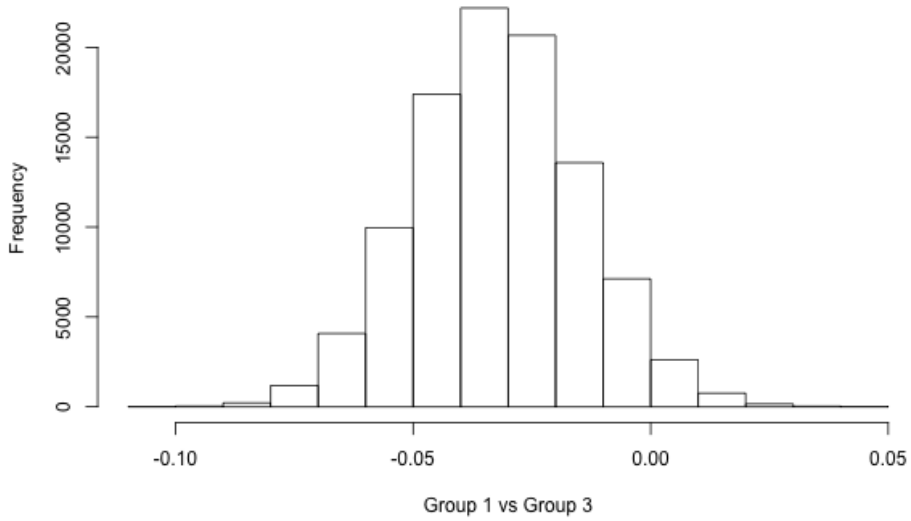


Figure 2. Samples from the posterior distribution of α_1 in the AUDIT-C model when comparing Group 2 versus Group 3 (Equation 1, back transformed). AUDIT-C: Alcohol Use Disorders Identification Test.

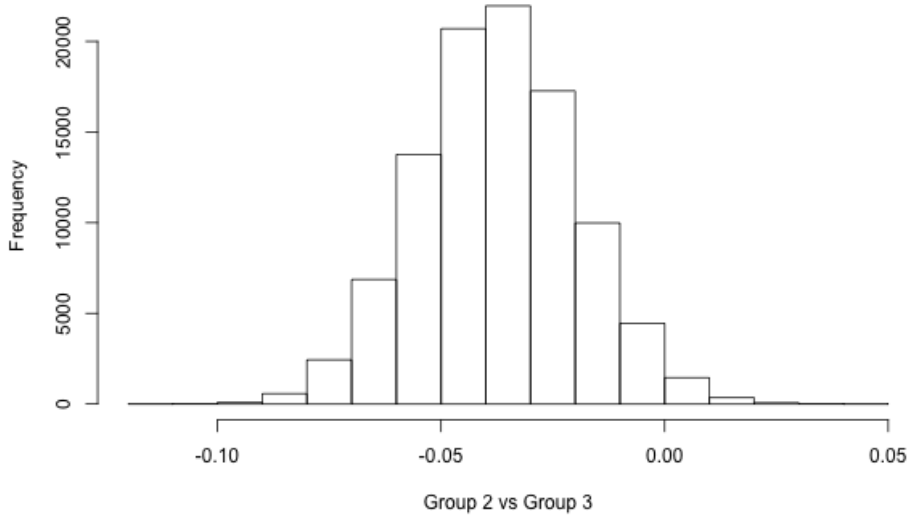


Figure 3. Samples from the posterior distribution of β_1 in the risky drinking model when comparing Group 1 versus Group 3 (Equation 2, exponentiated).

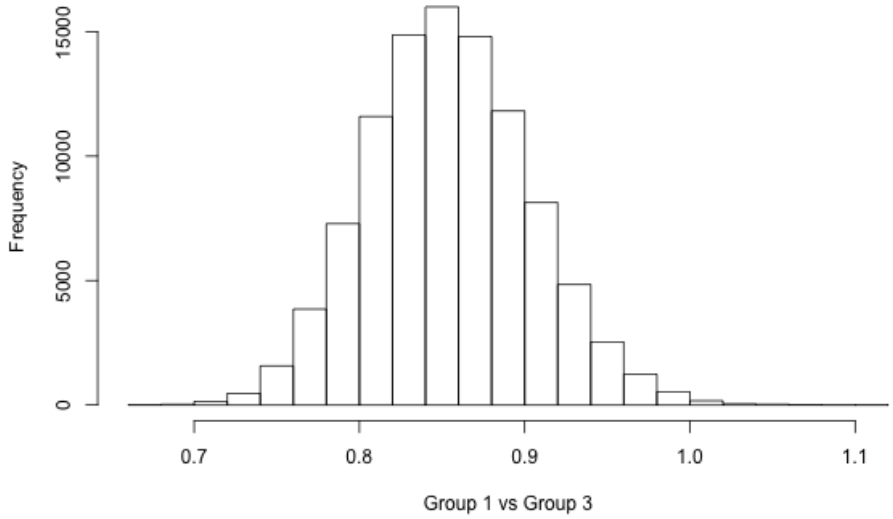


Figure 4. Samples from the posterior distribution of β_1 in the risky drinking model when comparing Group 2 versus Group 3 (Equation 2, exponentiated).

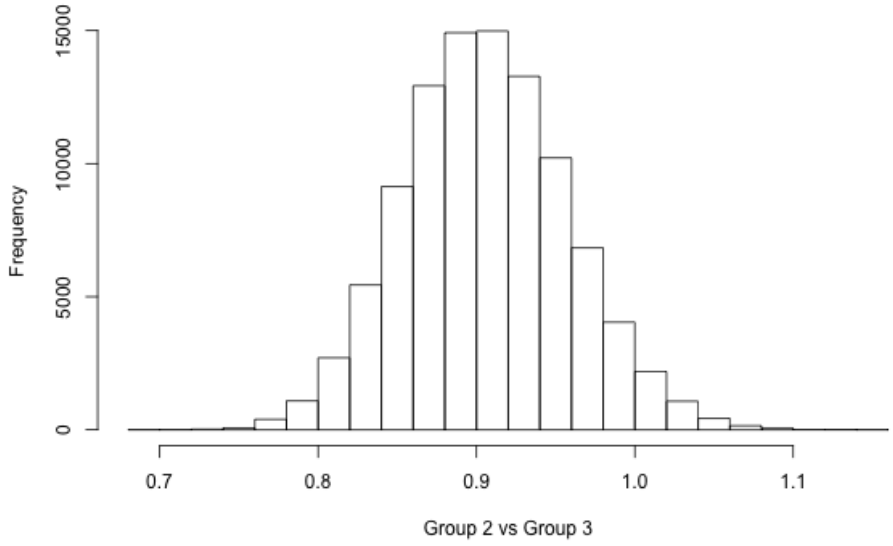


Table 2. Bayesian analysis of AUDIT-C at follow-up comparing Group 1 versus 3 and Group 2 versus 3.

	Group 1 versus 3			Group 2 versus 3		
	Threshold 1	Threshold 2	Threshold 3	Threshold 1	Threshold 2	Threshold 3
Regression coefficient[a] (AUDIT-C[b])	<0	<–0.02	<–0.04	<0	<–0.02	<–0.04
Marginal posterior probability (%)	96.4	75.7	32.9	98.1	83.7	44.4

[a]Back transformed linear regression coefficient (model adjusted for sex, age, university, and semester).

[b]AUDIT-C: Alcohol Use Disorders Identification Test.

Table 3. Bayesian analysis of risky drinking at follow-up comparing Group 1 versus 3 and Group 2 versus 3.

	Group 1 versus 3			Group 2 versus 3		
	Threshold 1	Threshold 2	Threshold 3	Threshold 1	Threshold 2	Threshold 3
Odds ratio[a] (Risky drinking)	<1	<0.9	<0.8	<1	<0.9	<0.8
Marginal posterior probability (%)	99.7	82.4	13.4	96.1	46.7	1.6

[a]Logistic regression coefficient in terms of odds ratios (model adjusted for sex, age, university, and semester).

Discussion

Key Findings

When comparing the analysis done in a null hypothesis framework with one done within the Bayesian framework, it is important to remind oneself of what the quantities represent as the questions being asked and answered are different.

The null hypothesis testing approach aims to put forth evidence about the population value of a parameter (ie, the existence of an effect on the entire population). The *P* value indicates how extreme the collected data are, given a fixed population value which is often set at a no-effect level. If the data is unlikely to have been generated from a population where the intervention has no effect, then the null hypothesis is rejected, and we can state that, with statistical significance, we believe that the intervention has a population effect. The fundamental issues with this approach have been discussed elsewhere [19-26], as has the problematic misinterpretation of *P* values and confidence intervals [27,28].

On the other hand, the Bayesian approach only concerns itself with the data at hand. It does not attempt to say anything about a population level effect, but instead calculates posterior distributions over model parameters. We can use these posterior distributions to calculate the probability of there being a difference between groups with respect to different trial outcomes.

Alcohol Use Disorders Identification Test

Null Hypothesis Framework

When contrasting AUDIT-C scores (Equation 1) using the null hypothesis framework (Table 1) we found no significant difference between Group 1 and Group 3 (*P*=.07), but there was

a significant difference between Group 2 and Group 3 (P=.04). This is somewhat counterintuitive, as Group 1 and 2 were given identical questionnaires, but Group 1 was also given feedback and advice. However, due to the nature of null hypothesis testing, we cannot discuss the effect of the feedback and advice component, since the very existence of an effect cannot be determined. Yet, we are to conclude that there does exist an effect with respect to responding to the questionnaire itself. Both P values are close to the conventional threshold of .05, and it is noteworthy that the entire discussion about effects would have changed had we adopted a .08 threshold or a .03 threshold.

In the original report [7], a direct comparison between Group 1 and Group 2 was also included but has been left out of this reanalysis for succinctness. The results were inconclusive, with no P values crossing the conventional threshold.

Bayesian Framework

The Bayesian approach (Table 2) suggests that the probability that Group 1 had a lower AUDIT-C score on average compared to Group 2 at follow-up was 96.4%, and when comparing Group 2 and Group 3 this probability was 98.1%. The model does not make a dichotomous decision about the existence of an effect but states the probability that a difference existed between the groups. Licensed by the randomization component of the trial design, we may conclude that this difference is due to the groups receiving different treatments. We may also conclude that the difference between groups is more likely than not to be greater than 0.02 units on the AUDIT-C scale but also is more likely than not to be less than 0.04.

Risky Drinking

Null Hypothesis Framework

In Table 1 we can see that the difference in prevalence of risky drinking between Group 1 and Group 3 was statistically significant, but not between Group 2 and Group 3. The existence of an effect on risky drinking is thus confirmed for the questionnaire plus feedback and advice intervention, but the evidence is inconclusive for the effect of the questionnaire alone. Recall that the situation was the opposite when analyzing AUDIT-C scores.

Bayesian Framework

The Bayesian approach (Table 3) suggests that there is a 99.7% probability that the prevalence of risky drinking was lower in Group 1 compared to Group 3, and that this probability was 96.1% when comparing Group 2 and Group 3. Note, however, that it was more likely than not that the OR was less than 0.9 when comparing Group 1 and Group 3, but not so when comparing Group 2 and Group 3. We can also see this when comparing Figure 3 and Figure 4, as most of the samples drawn when comparing Group 1 and Group 3 are to the left of 0.9, while they are centered around 0.9 when comparing Group 2 and Group 3. As was the case with AUDIT-C scores, we may attribute the difference between groups to the different treatments licensed by the trial design.

Clinical Significance

Clinically significant effect sizes are not universal, as they depend on the context in which the intervention can be offered and must be decided upon given cost, alternative interventions, ethical and practical concerns, and so on. One of the benefits of using a Bayesian approach is that we have access to a posterior distribution over the parameters of our model, which allows us to answer questions such as, “What is the probability that the effect is X or greater?” Therefore, we can evaluate the probability of clinically significant effect sizes in several different contexts. For instance, at the time of the AMADEUS-1 trial, student health care centers in Sweden did not have any means of reaching the entire student population with a brief intervention, thus there were no alternative interventions to the eSBI on trial. In addition, there was very little cost involved in adopting the eSBI into routine practice. Tables 1 and 3 indicate that there was a 4-percentage point difference in risky drinking between Groups 1 and 3 (OR<0.9; 82.4% probability), and this was considered a significant enough effect size to mandate a full-scale adoption of the intervention.

The years to come after the AMADEUS-1 trial saw many more trials of eSBIs, and as was mentioned earlier, meta-analyses suggest a small positive effect of eSBIs on the amount of alcohol consumed weekly in the short term.

Limitations

The AMADEUS-1 trial was unconventional in the sense that participants were randomized prior to being invited to the trial. This design allowed for a naturalistic study context and allowed for methodological advantages. However, participation rates were lower than would be expected in a more traditional setting where participants are randomized after registering interest in the trial (eg, only 36.2% [1798/4969] of participants allocated to Group 1 completed the eSBI). The overall follow-up rate was not remarkable at 52% (7764/14,910), which at the time was considered average for eHealth trials. Since missing at random cannot be guaranteed, effect sizes should be considered in the light that bias might have been introduced due to lower than ideal follow-up rates.

Summary

In the original publication of the AMADEUS-1 trial, we summarized the main results as follows [7]: There were consistently small differences in the anticipated direction in comparisons with group 3, which were possibly as a result of chance, with P<.10 for four of five comparisons for both groups 1 and 2.

Since not all P values from the hypotheses tests passed conventional thresholds, we could not conclude that the different groups had been affected by treating them differently. Unfortunately, the desire to dichotomize evidence prohibited any further discussion. This dichotomization of evidence creates issues when interpreting results from trials, as clearly the contradicting results when contrasting AUDIT-C scores is difficult to explain and communicate. Thus, we may ask if it is prudent for the student health care centers to decide to change their policy of offering eSBIs to all students since the .05

threshold was not broken? What if a different threshold was chosen? Should the results simply be discarded as inconclusive?

In this Bayesian reanalysis, we may instead summarize our findings as: There is 96.4% probability that Group 1 had a lower AUDIT-C score on average than did Group 3, and there is a 99.7% probability that the prevalence of risky drinking was lower in Group 1 compared to Group 3 (and a further 82.4% probability that the OR was less than 0.9). This then allows us to go forth and inspect the posterior distributions at effect sizes that are clinically significant in different contexts and discuss whether the intervention should be adopted into routine practice.

Conflicts of Interest

MB owns a private company that develops and distributes evidence-based lifestyle interventions to be used in health care settings, including student health care centers.

Multimedia Appendix 1
Trace plots.
[PDF File (Adobe PDF File), 397 KB - jmir_v21i11e14419_app1.pdf]

References

1. WHO, Management OSAT. Global Status Report on Alcohol and Health 2018. Geneva: WHO; 2018.
2. GBD 2016 Alcohol Collaborators. Alcohol use and burden for 195 countries and territories, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet 2018 Sep 22;392(10152):1015-1035 [FREE Full text] [doi: 10.1016/S0140-6736(18)31310-2] [Medline: 30146330]
3. Babor T, Caetano R, Casswell S, Edwards G, Giesbrecht N, Graham K, et al. Alcohol: No Ordinary Commodity: Research and Public Policy. Oxford: Oxford University Press; 2010.
4. McCambridge J, Bendtsen P, Bendtsen M, Nilsen P. Alcohol email assessment and feedback study dismantling effectiveness for university students (AMADEUS-1): study protocol for a randomized controlled trial. Trials 2012 Jul 06;13:49 [FREE Full text] [doi: 10.1186/1745-6215-13-49] [Medline: 22540638]
5. Bendtsen P, McCambridge J, Bendtsen M, Karlsson N, Nilsen P. Effectiveness of a proactive mail-based alcohol Internet intervention for university students: dismantling the assessment and feedback components in a randomized controlled trial. J Med Internet Res 2012 Oct 31;14(5):e142 [FREE Full text] [doi: 10.2196/jmir.2062] [Medline: 23113955]
6. McCambridge J, Bendtsen M, Karlsson N, White IR, Bendtsen P. Alcohol assessment & feedback by e-mail for university student hazardous and harmful drinkers: study protocol for the AMADEUS-2 randomised controlled trial. BMC Public Health 2013 Oct 10;13:949 [FREE Full text] [doi: 10.1186/1471-2458-13-949] [Medline: 24456668]
7. McCambridge J, Bendtsen M, Karlsson N, White IR, Nilsen P, Bendtsen P. Alcohol assessment and feedback by email for university students: main findings from a randomised controlled trial. Br J Psychiatry 2013 Nov;203(5):334-340 [FREE Full text] [doi: 10.1192/bjp.bp.113.128660] [Medline: 24072758]
8. Bendtsen P, Bendtsen M, Karlsson N, White IR, McCambridge J. Online Alcohol Assessment and Feedback for Hazardous and Harmful Drinkers: Findings From the AMADEUS-2 Randomized Controlled Trial of Routine Practice in Swedish Universities. J Med Internet Res 2015 Jul 09;17(7):e170 [FREE Full text] [doi: 10.2196/jmir.4020] [Medline: 26159179]
9. Rogers R. Cognitive and physiological processes in fear appeals and attitude change: A revised theory of protection motivation. In: Social Psychophysiology: A Sourcebook. New York City: Guilford Press; 1983:153-177.
10. Bandura A. Self-efficacy: The exercise of control. New York City: W H Freeman/Times Books/ Henry Holt & Co; 1997.
11. Ajzen I. The theory of planned behavior. Organizational Behavior and Human Decision Processes 1991 Dec;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-t]
12. Michie S, Whittington C, Hamoudi Z, Zarnani F, Tober G, West R. Identification of behaviour change techniques to reduce excessive alcohol consumption. Addiction 2012 Aug;107(8):1431-1440. [doi: 10.1111/j.1360-0443.2012.03845.x] [Medline: 22340523]
13. Garnett CV, Crane D, Brown J, Kaner EFS, Beyer FR, Muirhead CR, et al. Behavior Change Techniques Used in Digital Behavior Change Interventions to Reduce Excessive Alcohol Consumption: A Meta-regression. Ann Behav Med 2018 May 18;52(6):530-543 [FREE Full text] [doi: 10.1093/abm/kax029] [Medline: 29788261]
14. Smedslund G, Wollscheid S, Fang L, Nilsen W, Steiro A, Larun L. Effects of early, computerized brief interventions on risky alcohol use and risky cannabis use among young people. Campbell Systematic Reviews 2017 Apr 07;13(1):1-192. [doi: 10.4073/csr.2017.6]
15. Carey KB, Scott-Sheldon LAJ, Elliott JC, Garey L, Carey MP. Face-to-face versus computer-delivered alcohol interventions for college drinkers: a meta-analytic review, 1998 to 2010. Clin Psychol Rev 2012 Dec;32(8):690-703 [FREE Full text] [doi: 10.1016/j.cpr.2012.08.001] [Medline: 23022767]
16. Donoghue K, Patton R, Phillips T, Deluca P, Drummond C. The effectiveness of electronic screening and brief intervention for reducing levels of alcohol consumption: a systematic review and meta-analysis. J Med Internet Res 2014 Jun 02;16(6):e142 [FREE Full text] [doi: 10.2196/jmir.3193] [Medline: 24892426]

17. Bendtsen M. Electronic screening and brief intervention for hazardous and harmful drinkers among swedish university students: reanalysis of findings from a randomized controlled trial using a Bayesian framework (under review). J Med Internet Res 2019 Under review(forthcoming).
18. Bush K, Kivlahan DR, McDonell MB, Fihn SD, Bradley KA. The AUDIT alcohol consumption questions (AUDIT-C): an effective brief screening test for problem drinking. Ambulatory Care Quality Improvement Project (ACQUIP). Alcohol Use Disorders Identification Test. Arch Intern Med 1998 Sep 14;158(16):1789-1795. [doi: 10.1001/archinte.158.16.1789] [Medline: 9738608]
19. Bendtsen M. A Gentle Introduction to the Comparison Between Null Hypothesis Testing and Bayesian Analysis: Reanalysis of Two Randomized Controlled Trials. J Med Internet Res 2018 Oct 24;20(10):e10873 [FREE Full text] [doi: 10.2196/10873] [Medline: 30148453]
20. Amrhein V, Greenland S. Remove, rather than redefine, statistical significance. Nat Hum Behav 2018 Jan 25;2(1):4-4. [doi: 10.1038/s41562-017-0224-0] [Medline: 30980046]
21. Benjamin DJ, Berger JO, Johannesson M, Nosek BA, Wagenmakers E, Berk R, et al. Redefine statistical significance. Nat Hum Behav 2018 Jan 1;2(1):6-10. [doi: 10.1038/s41562-017-0189-z] [Medline: 30980045]
22. Berry D. A P-Value to Die For. Journal of the American Statistical Association 2017 Oct 30;112(519):895-897. [doi: 10.1080/01621459.2017.1316279]
23. Briggs WM. The Substitute for P-Values. Journal of the American Statistical Association 2017 Oct 30;112(519):897-898. [doi: 10.1080/01621459.2017.1311264]
24. Gelman A, Carlin J. Some Natural Solutions to the P-Value Communication Problem—and Why They Won't Work. Journal of the American Statistical Association 2017 Oct 30;112(519):899-901. [doi: 10.1080/01621459.2017.1311263]
25. Laber EB, Shedden K. Statistical Significance and the Dichotomization of Evidence: The Relevance of the ASA Statement on Statistical Significance and p-Values for Statisticians. J Am Stat Assoc 2017;112(519):902-904 [FREE Full text] [doi: 10.1080/01621459.2017.1311265] [Medline: 29348701]
26. Nuzzo R. Scientific method: statistical errors. Nature 2014 Feb 13;506(7487):150-152. [doi: 10.1038/506150a] [Medline: 24522584]
27. McShane BB, Gal D. Statistical Significance and the Dichotomization of Evidence. Journal of the American Statistical Association 2017 Oct 30;112(519):885-895. [doi: 10.1080/01621459.2017.1289846]
28. McShane BB, Gal D. Rejoinder: Statistical Significance and the Dichotomization of Evidence. Journal of the American Statistical Association 2017 Oct 30;112(519):904-908. [doi: 10.1080/01621459.2017.1323642]

Abbreviations

AMADEUS-1: Alcohol email assessment and feedback study dismantling effectiveness for university students
AUDIT-C: Alcohol Use Disorders Identification Test
eHealth: electronic health
eSBI: electronic screening and brief intervention
OR: odds ratio

Edited by G Eysenbach; submitted 17.04.19; peer-reviewed by D Gratzer, B Smith; comments to author 17.06.19; revised version received 06.08.19; accepted 14.08.19; published 07.11.19.

Please cite as:
Bendtsen M
Electronic Screening for Alcohol Use and Brief Intervention by Email for University Students: Reanalysis of Findings From a Randomized Controlled Trial Using a Bayesian Framework
J Med Internet Res 2019;21(11):e14419
URL: https://www.jmir.org/2019/11/e14419
doi:10.2196/14419
PMID:31697242

Corrigenda and Addenda

Correction: Predictors of Patients' Loyalty Toward Doctors on Web-Based Health Communities: Cross-Sectional Study

Tailai Wu[1], PhD; Zhuo Chen[2,3], PhD; Donglan Zhang[2], PhD; Xiang Wu[1], PhD; Ruoxi Wang[1], PhD

[1]School of Medicine and Health Management, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China
[2]Department of Health Policy and Management, College of Public Health, University of Georgia, Athens, GA, United States
[3]School of Economics, Faculty of Humanities and Social Sciences, University of Nottingham Ningbo China, Ningbo, China

Corresponding Author:
Xiang Wu, PhD
School of Medicine and Health Management
Tongji Medical College
Huazhong University of Science and Technology
13 Hangkong Road
Wuhan, 430030
China
Phone: 86 13071253919
Email: wuhsiang@hust.edu.cn

Related Article:

Correction of: https://www.jmir.org/2019/9/e14484/

(J Med Internet Res 2019;21(11):e16942) doi:10.2196/16942

In "Predictors of Patients' Loyalty Toward Doctors on Web-Based Health Communities: Cross-Sectional Study" by Wu et al (J Med Internet Res 2019;21(9):e14484), a minor error in the typesetting stage of publication resulted in the Acknowledgments section not being included in the final version of the article.

The Acknowledgments section has been added to the paper and appears as follows:

This work was partially supported by grants from the National Natural Science Foundation of China (No 71801100, 71671073, 71602063, 71704059, 71701075, and 71810107003).

The correction will appear in the online version of the paper on the JMIR website on November 7, 2019, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 06.11.19; this is a non–peer-reviewed article;accepted 06.11.19; published 07.11.19.

Please cite as:
Wu T, Chen Z, Zhang D, Wu X, Wang R
Correction: Predictors of Patients' Loyalty Toward Doctors on Web-Based Health Communities: Cross-Sectional Study
J Med Internet Res 2019;21(11):e16942
URL: http://www.jmir.org/2019/11/e16942/
doi:10.2196/16942
PMID:31697641

Corrigenda and Addenda

Title Correction: The Association Between Willingness of Frontline Care Providers' to Adaptively Use Telehealth Technology and Virtual Service Performance in Provider-to-Provider Communication: Quantitative Study

Hyeyoung Hah[1*], PhD; Deana Goldin[2*], PhD, DNP; Sejin Ha[3*], PhD

[1]Department of Information Systems and Business Analytics, Florida International University, Miami, FL, United States

[2]Nicole Wertheim College of Nursing & Health Sciences, Florida International University, Miami, FL, United States

[3]Retail, Hospitality, and Tourism Management, College of Education, Health, and Human Sciences, The University of Tennessee, Knoxville, Knoxville, TN, United States

[*]all authors contributed equally

Corresponding Author:
Hyeyoung Hah, PhD
Department of Information Systems and Business Analytics
Florida International University
11200 SW 8th Street
Miami, FL, 33199
United States
Phone: 1 3053484342
Email: hyeyoung.hah@gmail.com

Related Article:

Correction of: http://www.jmir.org/2019/8/e15087/

(J Med Internet Res 2019;21(11):e17123) doi:10.2196/17123

The title of this paper (J Med Internet Res 2019;21(8):e15087) has been changed from "The Association Between Willingness of Frontline Care Providers' to Adaptively Use of Telehealth Technology and Virtual Service Performance in Provider-to-Provider Communication: Quantitative Study" to "The Association Between Willingness of Frontline Care Providers' to Adaptively Use Telehealth Technology and Virtual Service Performance in Provider-to-Provider Communication: Quantitative Study".

The corrections will appear in the online version of the paper on the JMIR website on November 19, 2019, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article also has been resubmitted to those repositories.

Edited by N Duguay; submitted 19.11.19; this is a non–peer-reviewed article;accepted 19.11.19; published 19.11.19.

Please cite as:
Hah H, Goldin D, Ha S
Title Correction: The Association Between Willingness of Frontline Care Providers' to Adaptively Use Telehealth Technology and Virtual Service Performance in Provider-to-Provider Communication: Quantitative Study
J Med Internet Res 2019;21(11):e17123
URL: http://www.jmir.org/2019/11/e17123/
doi:10.2196/17123
PMID:31743890

Corrigenda and Addenda

Metadata Correction: A Virtual Counseling Application Using Artificial Intelligence for Communication Skills Training in Nursing Education: Development Study

Shefaly Shorey[1], PhD; Emily Ang[1], DNurs; John Yap[2], MA; Esperanza Debby Ng[1], BA; Siew Tiang Lau[1], PhD; Chee Kong Chui[3], PhD

[1]Alice Lee Centre for Nursing Studies, National University of Singapore, Singapore, Singapore

[2]Information Techonology, National University of Singapore, Singapore, Singapore

[3]Department of Mechanical Engineering, National University of Singapore, Singapore, Singapore

Corresponding Author:
Shefaly Shorey, PhD
Alice Lee Centre for Nursing Studies
National University of Singapore
Clinical Research Centre 10 Medical Drive
Singapore, 117597
Singapore
Phone: 65 66011294
Email: nurssh@nus.edu.sg

Related Article:

Correction of: https://www.jmir.org/2019/10/e14658

(J Med Internet Res 2019;21(11):e17064) doi:10.2196/17064

In “A Virtual Counseling Application Using Artificial Intelligence for Communication Skills Training in Nursing Education: Development Study” by Shorey et al (J Med Internet Res 2019;21(10):e14658), the affiliation of authors Shefaly Shorey and Emily Ang has been corrected from “Department of Mechanical Engineering, National University of Singapore” to “Alice Lee Centre for Nursing Studies, National University of Singapore”.

The contact information for Shefaly Shorey has also changed from “Department of Mechanical Engineering, National University of Singapore, Singapore, Singapore” to “Alice Lee Centre for Nursing Studies, National University of Singapore, Clinical Research Centre 10 Medical Drive, Singapore, 117597, Singapore”.

The correction will appear in the online version of the paper on the JMIR website on November 26, 2019, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 14.11.19; this is a non–peer-reviewed article;accepted 20.11.19; published 26.11.19.

Please cite as:
Shorey S, Ang E, Yap J, Ng ED, Lau ST, Chui CK
Metadata Correction: A Virtual Counseling Application Using Artificial Intelligence for Communication Skills Training in Nursing Education: Development Study
J Med Internet Res 2019;21(11):e17064
URL: http://www.jmir.org/2019/11/e17064/
doi:10.2196/17064
PMID:31770095

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org