<u>Viewpoint</u>

Evaluating Artificial Intelligence in Clinical Settings—Let Us Not Reinvent the Wheel

Kathrin Cresswell[1], PhD; Nicolette de Keizer[2,3], PhD; Farah Magrabi[4], PhD; Robin Williams[5], PhD; Michael Rigby[6], PhD; Mirela Prgomet[7], PhD; Polina Kukhareva[8], PhD; Zoie Shui-Yee Wong[9], PhD; Philip Scott[10], PhD; Catherine K Craven[11], PhD; Andrew Georgiou[4], PhD; Stephanie Medlock[2,12], PhD; Jytte Brender McNair[13], PhD; Elske Ammenwerth[14], PhD

[1]Usher Institute, The University of Edinburgh, Usher Building, Edinburgh, United Kingdom

[2]Amsterdam UMC, University of Amsterdam, Medical Informatics, Amsterdam, Netherlands

[3]Amsterdam Public Health Research Institute, Digital Health and Quality of Care, Amsterdam, Netherlands

[4]Australian Institute of Health Innovation, Macquarie University, Sydney, Australia

[5]Institute for the Study of Science, Technology and Innovation, The University of Edinburgh, Edinburgh, United Kingdom

[6]School of Social, Political and Global Studies and School of Primary, Community and Social Care, Keele University, Keele, United Kingdom

[7]Faculty of Medicine, Health and Human Sciences, Macquarie University, Sydney, Australia

[8]Department of Biomedical Informatics, University of Utah, Utah, UT, United States

[9]St. Luke's International University, Tokyo, Japan

[10]University of Wales Trinity St David, Swansea, United Kingdom

[11]University of Texas Health Science Center, San Antonio, TX, United States

[12]Amsterdam Public Health, Methodology & Aging & Later Life, Amsterdam, Netherlands

[13]Department of Health Science and Technology, Aalborg University, Aalborg, Denmark

[14]Institute of Medical Informatics, Private University for Health Sciences and Health Technology, UMIT TIROL, Hall in Tirol, Austria

Corresponding Author:
Kathrin Cresswell, PhD
Usher Institute
The University of Edinburgh
Usher Building
5-7 Little France Road
Edinburgh, EH16 4UX
United Kingdom
Phone: 44 131 650 6984
Email: <u>kathrin.cresswell@ed.ac.uk</u>

Abstract

Given the requirement to minimize the risks and maximize the benefits of technology applications in health care provision, there is an urgent need to incorporate theory-informed health IT (HIT) evaluation frameworks into existing and emerging guidelines for the evaluation of artificial intelligence (AI). Such frameworks can help developers, implementers, and strategic decision makers to build on experience and the existing empirical evidence base. We provide a pragmatic conceptual overview of selected concrete examples of how existing theory-informed HIT evaluation frameworks may be used to inform the safe development and implementation of AI in health care settings. The list is not exhaustive and is intended to illustrate applications in line with various stakeholder requirements. Existing HIT evaluation frameworks can help to inform AI-based development and implementation by supporting developers and strategic decision makers in considering relevant technology, user, and organizational dimensions. This can facilitate the design of technologies, their implementation in user and organizational settings, and the sustainability and scalability of technologies.

KEYWORDS

Introduction

The last two decades have seen rapid growth in artificial intelligence (AI) initiatives in health care settings, driven by the promises of improved treatment, quality, safety, and efficiency [1]. AI systems are computer algorithms that are able to mimic human intelligence to perform tasks. They are potentially capable of improving clinical decision-making. However, there is currently a lack of high-quality evidence of effectiveness, and an overoptimism regarding AI-based technologies in health care [2,3]. Many existing algorithms and applications fail to scale and migrate across settings [4], potentially leading to missed benefits or compromised patient safety.

Evidence from other sectors, such as finance and retail, may have limited applicability given the particular social, economic, technical processes, and legal challenges of health and social care settings [5]. Across the digital economy, AI has been successfully applied to historical data, for example, in financial forecasting [6] or retail marketing, where personalized advertisements have transformed consumer behavior [7]. These methods are harder to deploy in the more complex and sensitive settings of health and social care [5]. This is largely because developers and implementers focus on tool development and do not sufficiently draw on existing work to inform the conception and design of technologies, their use and optimization, and organizational strategies to implement them.

Theory-informed approaches to evaluation can help to ensure that technologies are effectively validated, implemented, and adopted. They can also help to ensure that systems do not result in unintended negative consequences, such as inappropriate or suboptimal care, exacerbated inequities, or clinician burnout [8]. Theories seek to explain complex relationships at an abstract level and can help to integrate a particular implementation with the empirical evidence base. As such, theory-informed evaluation frameworks can enable learning from experience, thus guiding developers, implementers, and evaluators through development, implementation, and optimization [9]. Ideally, the real-world experience gathered during this process is then used also to inform the refinement of evaluation frameworks.

Despite significant investments, there are currently only a few examples of the use of AI-based systems in health care and most systems are only beginning to be rolled out and embedded [10-12]. This is in contrast to the finance and retail sectors, where processes and products are standardized. To date, most activity has focused on diagnostic image-based systems and text or language processing, while complex precision medicine efforts are in very early stages of development. We here call for the increasing use of theory-informed approaches to evaluation to help ensure that developed systems can be adopted, scaled, and sustained within settings of use, and are safe and effective. Until now, this has not been done consistently, which has resulted in limited learning and limited ability to transfer learning across settings, as well as limited clinical and patient reassurance. If done appropriately, the implications for clinical settings are significant, as validated new knowledge can be disseminated and shared. This, in turn, obviates the need to learn through experience that can be painful, dangerous, and costly.

Unfortunately, despite increasing attention in research, the current application of theory-informed strategy and evaluation in AI practice is relatively limited in both health care and other sectors [13]. This may be due to a lack of understanding surrounding the theoretical literature (ie, why theories are useful in practice and how they may be used by different stakeholders), and the immediate focus of developers on demonstrating that technology works. Politically and managerially, there may be a drive to show modernization processes rather than making clinical and organizational decisions based on evidence-based outcomes. Where theories have been applied, these have been driven by business approaches to value creation in organizations [14], or by approaches designed to influence consumer behavior [15]. In these contexts, they have been strategically used to help address a particular stakeholder need (eg, how to maximize value through implementing AI in organizations and how to get consumers to accept AI technology). In health care, the range of stakeholders and associated needs however varies significantly from other sectors. While the managers and policymakers may focus on value and efficiency, patients are likely to be concerned about avoidable illness, and practitioners may focus on workloads and potential liability.

It is therefore often difficult to know what needs (and consequently what theory) to focus on and in what context. For example, while developers of technology now increasingly draw on cocreation with users to promote the adoption of AI, these approaches may not consider organizational drivers, workflow integration, multiplicity of stakeholders, or ethical considerations in implementation, thereby limiting the scalability of emerging applications.

Theory-informed approaches to evaluation in health care must be considered within their specific context, recognizing their relative positions and identifying which needs they address at various stages of the technology lifecycle. We aim to begin this journey by providing a conceptual overview of existing theory-informed frameworks that could usefully inform the development and implementation of AI-based technologies in health care. Despite some differences in technological properties and performance between AI- and non–AI-based technologies (Table 1) [16], many existing frameworks are likely to be applicable.

Table 1. Differences between artificial intelligence (AI)–based and non–AI-based health IT.

Applications	AI-based	Evidence	Non–AI-based	Evidence
Health services management	AI can help in optimizing resource allocation, scheduling, and workflow management by analyzing large data sets and identifying patterns and trends. For example, modeling of waiting times and underlying reasons	Limited evidence in relation to impact, mainly in relation to proof-of-concept [17,18]	Non–AI-based approaches typically rely on manual processes and human decision-making for resource management, scheduling, and workflow optimization. For example, patient flow management applications	High potential of data-driven approaches to improve organizational performance [19,20]
Predictive medicine	AI algorithms can analyze patient data, genetic information, and medical records to predict disease risks, treatment outcomes, and responses to therapies. This enables personalized medicine and targeted interventions	Many proof-of-concept studies but limited evidence in relation to how outputs are incorporated into clinical decision-making [21,22]	Non–AI-based approaches rely on statistical analysis and clinical expertise to make predictions about disease risks, treatment outcomes, and responses to therapies	Many proof-of-concept studies but limited evidence in relation to how outputs are incorporated into clinical decision-making [23,24]
Clinical decision support systems	AI to analyze large amounts of medical literature, patient data, and clinical guidelines to support clinical decision-making	Area of most focus, especially in imaging applications, AI has the potential to improve practitioner performance [25,26], but limited evidence surrounding organizational impacts or patient outcomes	Non–AI-based approaches rely on the expertise and experience of health care professionals, along with clinical guidelines and published research, to make clinical decisions	Demonstrated benefits for practitioner performance and patient outcomes in some areas of use (eg, drug-drug interactions) [27,28]
Laboratory and radiology information systems	Use of AI to detect abnormalities and to enhance the accuracy of diagnoses	Most progress has been made in relation to imaging [29,30], but limited attention has been paid to integration with organizational practices as above [26]	Non–AI-based diagnostics typically rely on visual inspection by health care professionals and manual analysis of patient data	This is associated with information overload but does take account of contextual factors
Patient data repositories	AI algorithms can process patient data to identify trends, patterns, and risk factors	Promising proof-of-concept studies, but limited implementation [31,32]	Patient data are stored in a centralized repository	Some evidence that digitized records and repositories can lead to improved quality, safety, and efficiency, but hard to assess and take a long time to materialize [33,34]
Population health management	Precision prevention approaches to identify populations at risk and tailor preventative interventions	Promising approaches to precision prevention in specific cohorts, but limited implementation [35,36]	Understanding factors that influence health outcomes and developing tailored interventions	Significant evidence of population health interventions [37,38]
Patient portals	AI-based symptom checkers and triage tools	Inconsistent evidence in relation to symptom checkers and triage tools, concerns in relation to diagnostic accuracy [39,40]	Access to generic informational resources	Tailored informational resources can improve satisfaction, involvement, and decision-making [41,42]
Telehealth and telecare	Online health assistants and chatbots	Mixed evidence of effectiveness usability and user satisfaction [43,44]	Access to generic informational resources.	Tailored informational resources can improve satisfaction, involvement, and decision-making [41,42]
Health information exchange	Extracting and converting unstructured or semistructured data into a standardized format	The use of free text data is still in its infancy but is promising. There is limited data on integration with existing ways of working and organizational functioning. [45,46]	Coding and transfer into standardized formats are often done by health care staff	Increased workloads for health care staff and coding are often not done accurately [47,48]

We here provide a conceptual overview of existing frameworks, focusing on practical applications of examples of existing theory-informed frameworks and their potential application to AI-based technologies in health care [49]. Frameworks were selected as examples illustrating these extracted categories. This work is not intended to be exhaustive but to provide a pragmatic introduction to the topic for nonspecialists [50,51].

To categorize frameworks in a meaningful way, we focused on their potential area of application and the particular interest or focus of various stakeholder groups who may need to draw on existing experience to inform their current efforts to develop, implement, and optimize AI-based technologies in health care settings.

XSL•FO

RenderX

Health IT Evaluation Frameworks and Their Potential Application to AI

Overview

The 3 distinct dimensions identified are illustrated in Table 2, along with potential applications of AI-based technologies in health care and example use cases. These include frameworks with a technology, user, and organizational focus. We discuss each of these categories, the application of exemplary frameworks, and practical implications for various stakeholders in the paragraphs below.

However, it is important to recognize that the categorization of frameworks provided here is a simplification. Various frameworks have common and, in some instances, overlapping elements. The categories presented are intended to facilitate navigation and application.

Table 2. Examples of the focus of existing health IT evaluation frameworks and their potential application to artificial intelligence.

Focus of the framework	Area of application	Example theoretical lenses	Practical implications	Stakeholders	Examples
Technology focus	• Informing the conception and design of technologies • To help AI[a] system developers design a system that is usable and useful within intended use settings	Human-centered design	Actively and iteratively involve end users in system design and development	End users and developers	A team had developed an algorithm to predict arterial fibrillation from electrocardiograms, but prospective users stated that the information would not change their practice [10]
User focus	• Informing and helping to optimize the use of technologies • To help developers and implementers understand the various contexts of use of AI as well as unintended consequences, and tailor systems to maximize benefits and minimize harms	Sociotechnical systems	Plan with users to effectively integrate the system in their work practices and monitor progress over time	End users, implementers	• IBM Watson encountered adoption-related issues, including usability and perceived usefulness of their oncology software, which eventually led to the abandonment • The system increased the workloads of doctors and made treatment recommendations that were viewed as unsafe by doctors [52]
Organizational focus	• Informing organizational strategies to implement technologies • To help AI system implementers integrate AI safely within existing organizational structures and processes	Institutional theory	Plan and monitor how systems and their outputs are integrated within and across organizational units and existing technological and social structures	End users, organizational stakeholders, and implementers	• Babylon Health UK (an AI-based remote service provider) failed because it did not fit with existing health system financing structures and cultures • Many patients from outside the local area enrolled in the service, which meant that the product was not commercially viable for local organizations [53]

[a]AI: artificial intelligence.

Frameworks With a Technology Focus

Many current AI applications in health care settings have been developed by AI specialists in laboratory settings. Consequently, they have struggled to successfully translate into clinical settings and deliver the performance achieved in research trials [54]. Frameworks with a technology focus can help to inform the "conception and design" of technologies, thereby helping to ensure that AI system developers design a system that is readily implemented and useful within intended use settings. For instance, techniques such as technology assessment and requirements analysis can help to identify use cases, constraints, and requirements that the new technology needs to fulfill.

Frameworks include, for example, design and usability frameworks such as the Health IT Usability Evaluation Model (Health-ITUEM) for evaluating mobile health technology [55]. This includes assessment of subjective properties of the technology from the perspective of users, which have been shown to be crucial to user adoption of technology, but that developers may not necessarily consider as a priority during the development process, including ease of use and perceived usefulness.

Frameworks With a User Focus

While use is crucial for the successful development of AI-based technology, empirical work has shown that systems may be used in ways other than intended, which may in turn result in unanticipated threats to organizational functioning and patient safety [56]. For example, users may develop workarounds to compensate for usability issues of technologies, but these workarounds may compromise the intended performance of a system [57]. Frameworks that focus on the user of the technology can help to address these issues and facilitate the "optimization of technology use". In doing so, they can help developers and implementers understand the various contexts of the use of AI-based technologies, as well as unintended consequences, and tailor systems to maximize benefits and minimize harms. For instance, a contextual analysis can help to gain a deep understanding of the various contexts in which a technology will be deployed. This includes examining cultural and social factors, as well as user behavior, user expectations, and existing systems or practices.

An example framework in this context is the Health Information Technology Evaluation Framework (HITREF), which includes an assessment of a technology's impact on quality of care as well as an assessment of unintended consequences [58].

Frameworks With an Organizational Focus

AI-based technologies are not adopted in a vacuum but must be integrated within organizational contexts. Previous work has shown that organizational strategies to implement health IT (HIT) and organizational cultures can have significant consequences for adoption and use [59]. For example, lack of integration with existing health information infrastructures can slow down system performance and impede practical use, and hence, impact adversely on safety and user experience [60]. Frameworks with an organizational focus can facilitate the development of "organizational strategies" to implement new technologies. In doing so, they can help AI system implementers integrate AI safely within existing organizational structures and processes. For instance, these can help to inform communication strategies, training programs, and support mechanisms to help users understand the benefits and risks of AI technologies and adapt to new roles and responsibilities.

An example of a framework with an organizational focus is the Safety Assurance Factors for Electronic Health Record Resilience (SAFER) guides, which help implementing organizations identify existing risks and facilitate the development of mitigation strategies to promote the effective integration of technologies within organizational processes [61].

Discussion

A range of theory-informed evaluation frameworks for diverse kinds of HIT already exist [62]. Although not all of these may be relevant for AI-based applications, many aspects of existing frameworks are likely to apply. Exploring the transferability of these dimensions, therefore, needs to be a central component of work going forward [63].

Existing frameworks examine various aspects of technology design, implementation, adoption, and optimization. On the most basic level, they can be distinguished according to their focus, which then influences their application and context of use. A simplified overview of selected HIT evaluation frameworks and their potential application to AI is shown in Table 2. Frameworks with a technology focus can help to inform the conception and design of technologies through actively and iteratively involving end users, bridging the gap between technology development and application. This can, in turn, mitigate risks around nonadoption due to a lack of need or actionable system outputs. Frameworks with a user focus can help to ensure that systems are effectively embedded with adoption contexts and thereby mitigate the risk of systems not being used or not being used as intended. Finally, frameworks with an organizational focus can help to ensure that systems fit with existing organizational structures, and thereby help to ensure sustained use over time and across contexts.

We recommend that researchers, implementers, and strategic decision makers consider the use of existing theory-informed HIT evaluation frameworks before embarking on an AI-related initiative. This can help to mitigate emerging risks and maximize the chances of successful implementation, adoption, and scaling. To achieve this, existing and emerging guidelines for the evaluation of AI must promote the use of theory-informed evaluation frameworks.

Although many of the frameworks are well-known in the academic clinical informatics community, there is an urgent need to incorporate them into general AI design, implementation, and evaluation activities, as they can help to facilitate learning from experience and ensure building on the existing empirical evidence base. Unfortunately, this is currently not routinely done, perhaps reflecting disciplinary silos leading to lessons having to be learned the hard way. This, in turn, potentially compromises the safety, quality, and sustainability of applications. For example, although AI applications in radiology are now getting more established, the existing evidence base focuses on demonstrating effectiveness in proof-of-concept or specific clinical settings (the technology dimension in Table 2) [25]. Wider organizational and user factors are somewhat neglected, potentially threatening the wider sustainability and acceptability of such applications.

Conclusions

We aimed to provide a conceptual overview of existing theory-informed frameworks that could usefully inform the development and implementation of AI-based technologies in health care, and we identified several frameworks with technological, user, and organizational foci. Future research could involve conducting a systematic review based on this pragmatic overview to synthesize existing evidence across evaluation frameworks, spanning the dimensions of technology, user, and organization.

Evaluation of AI-based systems needs to be based on theoretically informed empirical studies in contexts of implementation or use to ensure objectivity and rigor in establishing the benefits and thwarting risks. This will ensure that systems are based on relevant and transferable evidence and can be implemented safely and effectively. Theory-based HIT evaluation frameworks should be integrated into existing

and emerging guidelines for the evaluation of AI [64-66]. The examples of frameworks provided could also help to stimulate the development of other related frameworks that can guide further evaluation efforts.

Drawing effectively on theory-based HIT evaluation frameworks will help to strengthen the evidence-based implementation of AI systems in health care and help to refine and tailor existing theoretical approaches to AI-based HIT. Learning from the wealth of existing HIT evaluation experience will help patients, professionals, and wider health care systems.

Acknowledgments

Data Availability

Data sharing is not applicable to this article as no data sets were generated or analyzed during this study.

Authors' Contributions

KC led on drafting of the manuscript and all authors (NDK, FM, RW, MR, MP, PK, ZSW, PS, CKC, AG, SM, JBM, and EA) critically commented on various iterations.

Conflicts of Interest

None declared.

References

1. Bates DW, Levine D, Syrowatka A, Kuznetsova M, Craig KJT, Rui A, et al. The potential of artificial intelligence to improve patient safety: a scoping review. NPJ Digit Med. 2021;4(1):1-8. [FREE Full text] [doi: 10.1038/s41746-021-00423-6] [Medline: 33742085]
2. Magrabi F, Ammenwerth E, McNair JB, De Keizer NF, Hyppönen H, Nykänen P, et al. Artificial intelligence in clinical decision support: challenges for evaluating AI and practical implications. Yearb Med Inform. 2019;28(01):128-134. [doi: 10.1055/s-0039-1677903]
3. Patel VL, Kannampallil. Cognitive informatics in biomedicine and healthcare. J Biomed Inform. 2015;53:3-14. [doi: 10.1016/j.jbi.2014.12.007] [Medline: 25541081]
4. Lobo JL, Del Ser J, Bifet A, Kasabov N. Spiking neural networks and online learning: an overview and perspectives. Neural Netw. 2020;121:88-100. [doi: 10.1016/j.neunet.2019.09.004] [Medline: 31536902]
5. Yu KH, Beam AL, Kohane IS. Artificial intelligence in healthcare. Nat Biomed Eng. 2018;2(10):719-731. [doi: 10.1038/s41551-018-0305-z] [Medline: 31015651]
6. Li X, Sigov A, Ratkin L, Ivanov LA, Li L. Artificial intelligence applications in finance: a survey. Journal of Management Analytics. 2023;10(4):676-692. [doi: 10.1080/23270012.2023.2244503]
7. Weber FD, Schütte R. State-of-the-art and adoption of artificial intelligence in retailing. DPRG. 2019;21(3):264-279. [doi: 10.1108/dprg-09-2018-0050]
8. Wiegand T, Krishnamurthy R, Kuglitsch M, Lee N, Pujari S, Salathé M, et al. WHO and ITU establish benchmarking process for artificial intelligence in health. Lancet. 2019;394(10192):9-11. [doi: 10.1016/S0140-6736(19)30762-7] [Medline: 30935732]
9. Craven CK, Doebbeling B, Furniss D, Holden RJ, Lau F, Novak LL. Evidence-based Health Informatics Frameworks for Applied Use. Stud Health Technol Inform. 2016;222:77-89. [Medline: 27198094]
10. Williams R, Anderson S, Cresswell K, Kannelønning MS, Mozaffar H, Yang X. Domesticating AI in medical diagnosis. Technology in Society. 2024:102469. [doi: 10.1016/j.techsoc.2024.102469]
11. Fleuren LM, Thoral P, Shillan D, Ercole A, Elbers PWG, Right Data Right Now Collaborators. Machine learning in intensive care medicine: ready for take-off? Intensive Care Med. 2020;46(7):1486-1488. [doi: 10.1007/s00134-020-06045-y] [Medline: 32399747]
12. Mackenzie SC, Sainsbury CAR, Wake DJ. Diabetes and artificial intelligence beyond the closed loop: a review of the landscape, promise and challenges. Diabetologia. 2024;67(2):223-235. [FREE Full text] [doi: 10.1007/s00125-023-06038-8] [Medline: 37979006]
13. Reim W, Åström J, Eriksson O. Implementation of artificial intelligence (AI): a roadmap for business model innovation. AI. 2020;1(2):180-191. [doi: 10.3390/ai1020011]
14. Enholm IM, Papagiannidis E, Mikalef P, Krogstie J. Artificial intelligence and business value: a literature review. Inf Syst Front. 2021;24:1709-1734. [FREE Full text] [doi: 10.1007/s10796-021-10186-w]

15. Vlačić B, Corbo L, Costa e Silva S, Dabić M. The evolving role of artificial intelligence in marketing: A review and research agenda. Journal of Business Research. 2021;128:187-203. [doi: 10.1016/j.jbusres.2021.01.055]

16. Secinaro S, Calandra D, Secinaro A, Muthurangu V, Biancone P. The role of artificial intelligence in healthcare: a structured literature review. BMC Med Inform Decis Mak. 2021;21(1):125. [FREE Full text] [doi: 10.1186/s12911-021-01488-9] [Medline: 33836752]

17. Pianykh OS, Guitron S, Parke D, Zhang C, Pandharipande P, Brink J, et al. Improving healthcare operations management with machine learning. Nat Mach Intell. 2020;2(5):266-273. [doi: 10.1038/s42256-020-0176-3]

18. Sebla AK. Use of artificial intelligence in health services management in Türkiye. International Journal of Health Services Research and Policy. 2023;8(2):139-161. [doi: 10.33457/ijhsrp.1298068]

19. Madsen LB. Data-Driven Healthcare: How Analytics and BI are Transforming the Industry. Hoboken, NJ. John Wiley & Sons; 2014.

20. Enticott J, Johnson A, Teede H. Learning health systems using data to drive healthcare improvement and impact: a systematic review. BMC Health Serv Res. 2021;21(1):200. [FREE Full text] [doi: 10.1186/s12913-021-06215-8] [Medline: 33663508]

21. Collins GS, Moons KGM. Reporting of artificial intelligence prediction models. Lancet. 2019;393(10181):1577-1579. [doi: 10.1016/S0140-6736(19)30037-6] [Medline: 31007185]

22. Schork NJ. Artificial intelligence and personalized medicine. Cancer Treat Res. 2019;178:265-283. [FREE Full text] [doi: 10.1007/978-3-030-16391-4_11] [Medline: 31209850]

23. Bellazzi R, Ferrazzi F, Sacchi L. Predictive data mining in clinical medicine: a focus on selected methods and applications. Wiley Interdiscip Rev Data Min Knowl Discov. 2011;1(5):416-430. [doi: 10.1002/widm.23]

24. Kent DM, Steyerberg E, van Klaveren D. Personalized evidence based medicine: predictive approaches to heterogeneous treatment effects. BMJ. 2018;363:k4245. [FREE Full text] [doi: 10.1136/bmj.k4245] [Medline: 30530757]

25. Kelly BS, Judge C, Bollard SM, Clifford SM, Healy GM, Aziz A, et al. Radiology artificial intelligence: a systematic review and evaluation of methods (RAISE). Eur Radiol. 2022;32(11):7998-8007. [FREE Full text] [doi: 10.1007/s00330-022-08784-6] [Medline: 35420305]

26. Farič N, Hinder S, Williams R, Ramaesh R, Bernabeu MO, van Beek E, et al. Early experiences of integrating an artificial intelligence-based diagnostic decision support system into radiology settings: a qualitative study. J Am Med Inform Assoc. 2023;31(1):24-34. [FREE Full text] [doi: 10.1093/jamia/ocad191] [Medline: 37748456]

27. Bright TJ, Wong A, Dhurjati R, Bristow E, Bastian L, Coeytaux RR, et al. Effect of clinical decision-support systems: a systematic review. Ann Intern Med. 2012;157(1):29-43. [FREE Full text] [doi: 10.7326/0003-4819-157-1-201207030-00450] [Medline: 22751758]

28. Greenes R, Del FG, editors. Clinical Decision Support and Beyond Progress and Opportunities in Knowledge-Enhanced Health and Healthcare. Cambridge, MA. Academic Press; 2023.

29. Hafizović L, Čaušević A, Deumić A, Spahić Bećirović L, Gurbeta Pokvić L, Badnjević A. The use of artificial intelligence in diagnostic medical imaging: systematic literature review. 2021. Presented at: 2021 IEEE 21st International Conference on Bioinformatics and Bioengineering (BIBE); 15 December 2021:1-6; Kragujevac, Serbia. [doi: 10.1109/bibe52308.2021.9635307]

30. Rebitzer JB, Rege M, Shepard C. Influence, information overload, and information technology in health care. Adv Health Econ Health Serv Res. 2008;19:43-69. [Medline: 19548513]

31. Chen X, Liu Z, Wei L, Yan J, Hao T, Ding R. A comparative quantitative study of utilizing artificial intelligence on electronic health records in the USA and China during 2008-2017. BMC Med Inform Decis Mak. 2018;18(Suppl 5):117. [FREE Full text] [doi: 10.1186/s12911-018-0692-9] [Medline: 30526643]

32. Sarwar T, Seifollahi S, Chan J, Zhang X, Aksakalli V, Hudson I, et al. The secondary use of electronic health records for data mining: data characteristics and challenges. ACM Comput. Surv. 2022;55(2):1-40. [doi: 10.1145/3490234]

33. 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;8(1):e1000387. [FREE Full text] [doi: 10.1371/journal.pmed.1000387] [Medline: 21267058]

34. Bergmo TS. How to measure costs and benefits of eHealth interventions: an overview of methods and frameworks. J Med Internet Res. 2015;17(11):e254. [FREE Full text] [doi: 10.2196/jmir.4521] [Medline: 26552360]

35. Loomans-Kropp HA, Umar A. Cancer prevention and screening: the next step in the era of precision medicine. NPJ Precis Oncol. 2019;3:3. [FREE Full text] [doi: 10.1038/s41698-018-0075-9] [Medline: 30701196]

36. Herman WH, Ye W. Precision prevention of diabetes. Diabetes Care. 2023;46(11):1894-1896. [doi: 10.2337/dci23-0052] [Medline: 37890107]

37. Kaplan RM. Two pathways to prevention. American Psychologist. 2000;55(4):382-396. [doi: 10.1037//0003-066x.55.4.382]

38. Jacka FN, Mykletun A, Berk M. Moving towards a population health approach to the primary prevention of common mental disorders. BMC Med. 2012;10:149. [FREE Full text] [doi: 10.1186/1741-7015-10-149] [Medline: 23186355]

39. Riboli-Sasco E, El-Osta A, Alaa A, Webber I, Karki M, El Asmar ML, et al. Triage and diagnostic accuracy of online symptom checkers: systematic review. J Med Internet Res. 2023;25:e43803. [FREE Full text] [doi: 10.2196/43803] [Medline: 37266983]

XSL•FO
RenderX

40. Ilicki J. Challenges in evaluating the accuracy of AI-containing digital triage systems: a systematic review. PLoS One. 2022;17(12):e0279636. [FREE Full text] [doi: 10.1371/journal.pone.0279636] [Medline: 36574438]

41. Reynolds TL, Ali N, McGregor E, O'Brien T, Longhurst C, Rosenberg AL, et al. Understanding patient questions about their medical records in an online health forum: opportunity for patient portal design. American Medical Informatics Association; 2017. Presented at: AMIA annual symposium proceedings; 16 April 2018:1468; Washington, DC.

42. Ely JW, Osheroff JA, Ebell MH, Chambliss ML, Vinson DC, Stevermer JJ, et al. Obstacles to answering doctors' questions about patient care with evidence: qualitative study. BMJ. 2002;324(7339):710. [FREE Full text] [doi: 10.1136/bmj.324.7339.710] [Medline: 11909789]

43. Milne-Ives M, de Cock C, Lim E, Shehadeh MH, de Pennington N, Mole G, et al. The effectiveness of artificial intelligence conversational agents in health care: systematic review. J Med Internet Res. 2020;22(10):e20346. [FREE Full text] [doi: 10.2196/20346] [Medline: 33090118]

44. Wilson L, Marasoiu M. The development and use of chatbots in public health: scoping review. JMIR Hum Factors. 2022;9(4):e35882. [FREE Full text] [doi: 10.2196/35882] [Medline: 36197708]

45. Savova GK, Masanz JJ, Ogren PV, Zheng J, Sohn S, Kipper-Schuler KC, et al. Mayo clinical text analysis and knowledge extraction system (cTAKES): architecture, component evaluation and applications. J Am Med Inform Assoc. 2010;17(5):507-513. [FREE Full text] [doi: 10.1136/jamia.2009.001560] [Medline: 20819853]

46. Li D, Simon G, Chute CG, Pathak J. Using association rule mining for phenotype extraction from electronic health records. AMIA Jt Summits Transl Sci Proc. 2013;2013:142-146. [FREE Full text] [Medline: 24303254]

47. Bajaj Y, Crabtree J, Tucker AG. Clinical coding: how accurately is it done? Clinical Governance: An International Journal. 2007;12(3):159-169. [doi: 10.1108/14777270710775873]

48. Campbell S, Giadresco K. Computer-assisted clinical coding: a narrative review of the literature on its benefits, limitations, implementation and impact on clinical coding professionals. Health Inf Manag. 2020;49(1):5-18. [doi: 10.1177/1833358319851305] [Medline: 31159578]

49. Grant MJ, Booth A. A typology of reviews: an analysis of 14 review types and associated methodologies. Health Info Libr J. 2009;26(2):91-108. [FREE Full text] [doi: 10.1111/j.1471-1842.2009.00848.x] [Medline: 19490148]

50. Salkind NJ, editor. Encyclopedia of Research Design. North America. Sage; 2010.

51. Glasgow RE. What does it mean to be pragmatic? Pragmatic methods, measures, and models to facilitate research translation. Health Educ Behav. 2013;40(3):257-265. [doi: 10.1177/1090198113486805] [Medline: 23709579]

52. Faheem H, Dutta S. Artificial intelligence failure at IBM' Watson for Oncology'. IUP Journal of Knowledge Management. 2023;21(3):47-75.

53. Mahase E. Babylon looks to sell GP at hand and other UK business amid financial issues. BMJ. 2023;382:1835. [doi: 10.1136/bmj.p1835] [Medline: 37558246]

54. Ben-Israel D, Jacobs WB, Casha S, Lang S, Ryu WHA, de Lotbiniere-Bassett M, et al. The impact of machine learning on patient care: a systematic review. Artif Intell Med. 2020;103:101785. [doi: 10.1016/j.artmed.2019.101785] [Medline: 32143792]

55. Brown W, Yen PY, Rojas M, Schnall R. Assessment of the health IT usability evaluation model (Health-ITUEM) for evaluating mobile health (mHealth) technology. J Biomed Inform. 2013;46(6):1080-1087. [FREE Full text] [doi: 10.1016/j.jbi.2013.08.001] [Medline: 23973872]

56. Ash JS, Berg M, Coiera E. Some unintended consequences of information technology in health care: the nature of patient care information system-related errors. J Am Med Inform Assoc. 2004;11(2):104-112. [FREE Full text] [doi: 10.1197/jamia.M1471] [Medline: 14633936]

57. Cresswell KM, Mozaffar H, Lee L, Williams R, Sheikh A. Workarounds to hospital electronic prescribing systems: a qualitative study in english hospitals. BMJ Qual Saf. 2017;26(7):542-551. [FREE Full text] [doi: 10.1136/bmjqs-2015-005149] [Medline: 27129493]

58. Sockolow PS, Bowles KH, Rogers ML. Health information technology evaluation framework (HITREF) comprehensiveness as assessed in electronic point-of-care documentation systems evaluations. Stud Health Technol Inform. 2015;216:406-409. [Medline: 26262081]

59. Cresswell K, Sheikh A. Organizational issues in the implementation and adoption of health information technology innovations: an interpretative review. Int J Med Inform. 2013;82(5):e73-e86. [doi: 10.1016/j.ijmedinf.2012.10.007] [Medline: 23146626]

60. Cresswell KM, Mozaffar H, Lee L, Williams R, Sheikh A. Safety risks associated with the lack of integration and interfacing of hospital health information technologies: a qualitative study of hospital electronic prescribing systems in England. BMJ Qual Saf. 2017;26(7):530-541. [FREE Full text] [doi: 10.1136/bmjqs-2015-004925] [Medline: 27037303]

61. Sittig DF, Ash JS, Singh H. The SAFER guides: empowering organizations to improve the safety and effectiveness of electronic health records. Am J Manag Care. 2014;20(5):418-423. [FREE Full text] [Medline: 25181570]

62. Scott P, de KN, Georgiou A, editors. Applied Interdisciplinary Theory in Health Informatics: A Knowledge Base for Practitioners. Netherlands. IOS Press; 2019.

63. Schloemer T, Schröder-Bäck P. Criteria for evaluating transferability of health interventions: a systematic review and thematic synthesis. Implement Sci. 2018;13(1):88. [FREE Full text] [doi: 10.1186/s13012-018-0751-8] [Medline: 29941011]

XSL•FO

RenderX

64. Nykänen P, Brender J, Talmon J, de Keizer N, Rigby M, Beuscart-Zephir MC, et al. Guideline for good evaluation practice in health informatics (GEP-HI). Int J Med Inform. 2011;80(12):815-827. [doi: 10.1016/j.ijmedinf.2011.08.004] [Medline: 21920809]

65. Talmon J, Ammenwerth E, Brender J, de Keizer N, Nykänen P, Rigby M. STARE-HI -statement on reporting of evaluation studies in health informatics. Yearb Med Inform. 2009;78(1):23-31. [Medline: 19855867]

66. Vasey B, Nagendran M, Campbell B, Clifton DA, Collins GS, Denaxas S, et al. DECIDE-AI expert group. Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. Nat Med. 2022;28(5):924-933. [doi: 10.1038/s41591-022-01772-9] [Medline: 35585198]

Abbreviations

AI: artificial intelligence
Health-ITUEM: Health IT Usability Evaluation Model
HIT: health IT
HITREF: Health Information Technology Evaluation Framework
SAFER: Safety Assurance Factors for Electronic Health Record Resilience