Viewpoint

Ethical Imperatives for Working With Diverse Populations in Digital Research

Jonathan Herington[1], PhD; Kay Connelly[2,3], PhD; Judy Illes[4], PhD

[1]Department of Health Humanities and Bioethics, University of Rochester, Rochester, NY, United States

[2]Department of Informatics, Indiana University, Bloomington, IN, United States

[3]College of Engineering, Michigan State University, Lansing, MI, United States

[4]Neuroethics Canada, Division of Neurology, Department of Medicine, University of British Columbia, Vancouver, BC, Canada

Corresponding Author:
Jonathan Herington, PhD
Department of Health Humanities and Bioethics
University of Rochester
Box 676
601 Elmwood Ave
Rochester, NY, 14642
United States
Phone: 1 (585) 275 5800
Fax: 1 (585) 276 0152
Email: jonathan.herington@rochester.edu

Abstract

Digital research methodologies are driving a revolution in health technology but do not yet fully engage diverse and historically underrepresented populations. In this paper, we explore the ethical imperative for such engagement alongside accompanying challenges related to recruitment, appreciation of risk, and confidentiality, among others. We critically analyze existing research ethics frameworks and find that their reliance on individualistic and autonomy-focused models of research ethics does not offer adequate protection in the context of the diversity imperative. To meet the requirements of justice and inclusivity in digital research, methods will benefit from a reorientation toward more participatory practices.

Introduction

In the past 2 decades, there has been a rapid increase in human subject research that uses digital research methodologies. We define digital research as human subjects research through which data are collected by electronic devices such as laptops, mobile devices, wearables, pervasive sensors, digital cameras, and other emerging technologies. This includes research that involves the direct collection of digital data for research only, the parallel collection of data for service provision with research as an ancillary purpose, as well as so-called secondary research that repurposes data (eg, from a biometric watch) originally collected for nonresearch purposes. These digital research methods often yield data used to train, test, and validate machine learning models, and thus are at the heart of questions about the accuracy, transparency, and fairness of machine learning systems in health care and elsewhere [1-4]. Calls to diversify the data sets upon which machine learning systems are trained imply a call to diversify the populations that participate in digital research.

In this paper, we explore the ethical dimensions of conducting digital research with and for diverse research populations. To begin, we make an important distinction between marginalized and diverse populations. We define marginalized populations as research populations or subpopulations that are wholly or mostly comprised of people from communities that have historically been exposed to special risks such as socioeconomic or health threats or to which researchers have special obligations, owing to differential power dynamics or historical instances of research exploitation. We define diverse research populations as those that include participants of different ages, genders, races, ethnicities, religions, incomes, literacies, educational backgrounds, languages, cultural norms, and disabilities. Different kinds of diversity will be relevant in different contexts. Broadly representative research populations include many

different kinds of diverse subpopulations with sufficient participants from each population so as to proportionately reflect the demographics of the overall population, even if it is impossible to capture all attributes without a very large study population.

Prior work in research ethics has often focused on the special ethical risks and responsibilities that arise in the conduct of research with people who are members of marginalized populations or otherwise vulnerable. For example, genomics research with indigenous communities has long been held to pose special risks to individuals and communities alike and requires enhanced protocols for collaboration, cultural competency, transparency, and community capacity building [5]. Neuroscience research has followed suit as perspectives on the brain and mind extend well beyond western concepts [6,7], and equitable access to translational interventions remains unrealized [8-10]. Collaborative approaches to research, including participatory action methods toward the cocreation of studies and knowledge sharing have been developed in response. The structured guidance offered by The First Nations principles of "ownership, control, access, and possession" (commonly known as OCAP) and the Collective benefit, Authority to control, Responsibility, and Ethics (CARE) Principles of Indigenous Data Governance of the Global Indigenous Data Alliance [11,12] are 2 examples. Likewise, research on health monitoring tools for patients with rare diseases requires balancing their privacy and unique needs [13].

In this paper, we explore the suitability of conventional human subject protections [14] for dealing with the special challenges of research populations with a wide range of interests, cultures, and capacities. Respect for persons, beneficence, and justice have always been basic principles of research with human subjects [14], but the latter, long neglected among the 3, must now become an equal imperative [15-17]. We begin by revisiting the ethical justification for creating diverse and broadly representative research populations in digital research. Next, we identify the different kinds of ethical risks that researchers must address in the context of digital research, including diverse capacities, cultures, risks, and understandings of technology. Finally, we critically analyze the capacity of existing research ethics frameworks to deal with these issues and provide recommendations for improving the capacity of digital researchers to build just and sustainable relationships with diverse research populations. We conclude that the inclusion of marginalized communities is a facet of justice that is especially heightened in the context of digital research, where the range of experiences with, access to, and meaningfulness of digital technologies varies widely [18].

Imperative for Diverse Research Populations

The ethical responsibility to include diverse research populations in digital research is grounded in a commitment to ensure both the social value of research is widely realized [19,20] and that there is a fair distribution of the burdens and benefits of the research [16]. This obligation attaches to all research, but in the context of digital research, there are at least 3 reasons why realizing these values require a special emphasis on diverse populations.

First, digital research fuels the production of algorithmic tools that have long been accused of biases that differentially distribute opportunities and harms in diverse populations [21-23]. Many of these harms are due to sampling bias in training sets [24]. Others are due to differential measurement bias for some subpopulations, where methods of collection or proxy variables produce more measurement error for some groups than others [2,25]. A prominent example of this is bias in the accuracy of wearables for users with dark skin tones [26,27]. Carefully diversifying data sets in a way that fairly represents the underlying constructs within every subpopulation is thus crucial to remediating these biases in algorithmic tools.

Second, digital research often relies upon convenience samples or secondary data sets that capture existing users of the collection technology. This has the effect of skewing data toward individuals who are White, wealthy, young, and healthy. For instance, studies of patterns of bicycle trail usage may use deidentified data from the Strava app (Strava, Inc), which does not accurately represent the general population or even the population of cyclists [28]. The use of preexisting users of digital technologies as the recruitment pool or data set allows for the propagation of disparities in access and uptake of digital technologies. For instance, there are well-established racial imbalances in the uptake of telemedicine [29,30], eHealth record usage [31], and wearables [28]. Indeed, there are deep imbalances by age, rural and urban residency, and socioeconomic status in access to key resources for digital research such as broadband internet [32,33]. This imbalance is even more pronounced when doing research globally [34].

Third, well-conducted digital research can improve the suitability and usability of these technologies for underserved groups. Digital technologies have the capacity to ameliorate or compensate for existing racial, geographic, and socioeconomic disparities. For instance, appropriately deployed and supported telemedicine can improve the management of chronic disease and access to specialist consultations for people in rural or low-mobility populations [9,35]. Efforts to ameliorate racial bias in the accuracy of wearables such as pulse-oximeter accuracy, for example, require recruiting a diverse population during test and validation research [26].

Existing Frameworks

In the foundational text of research ethics in North America [14], the principle of justice primarily focused on fair subject selection. This focus emphasizes both that individuals are ncluded regardless of race and gender and that an effort is made to balance the burden and benefits of research at the population level [14]:

> *… even if individual researchers are treating their research subjects fairly, and even if IRBs are taking care to assure that subjects are selected fairly within a particular institution, unjust social patterns may nevertheless appear in the overall distribution of the burdens and benefits of research.*

However, the principles of respect for persons and beneficence also acknowledge, at least implicitly, that fairly sharing the burden and benefits of research among the members of a diverse society presents challenges. Respect for persons requires that participants in research are treated as individuals, not simply as members of demographic groups, and that people with diminished autonomy due to cognitive, physical, or social impediments are afforded extra protection. Beneficence requires that the balance of risks and benefits of participation for each individual subject—including vulnerable subjects—is favorable. The Belmont principles thus deal with the difficulties of diversity in research populations by requiring individualized assessment of consent, risks, and benefits for each subject.

In the US context, the Belmont framework is expressed in regulation through the Common Rule [36]. This regulatory framework enshrines additional protections for 3 existing special populations: children, pregnant persons, and prisoners. In addition, it requires researchers to consider the status of any participants who may be "vulnerable to coercion and undue influence," including military personnel, people who are cognitively disabled, elderly persons, ethnic minorities, refugees, and people who are socioeconomically disadvantaged. While the Common Rule does not specifically advocate for the inclusion of First Nation peoples within research oversight, it does recommend researchers comply with tribal sovereignty, including community-operated institutional review boards [37].

In Canada, the regulatory framework for research is encapsulated in the Tri-Council Policy Statement [38]. Of note, the policy statement explicitly requires an appropriate inclusion of diverse research subjects [38]:

> *Taking into account the scope and objectives of their research, researchers should be inclusive in selecting participants. Researchers shall not exclude individuals from the opportunity to participate in research on the basis of attributes such as culture, language, religion, race, disability, sexual orientation, ethnicity, linguistic proficiency, gender or age, unless there is a valid reason for the exclusion.*

Moreover, there has been a much greater effort to incentivize the careful inclusion of First Nations peoples in research [38]. This is complemented by the OCAP© framework mentioned earlier.

Challenges of Diversity

Overview

Broadly representative research populations will be composed of a range of individuals with different capacities, interests, values, and social positions. This poses a problem for attempts to assess the risks and benefits of a particular research design when researchers cannot individualize risk assessment. In particular, researchers cannot assume that the risks and benefits of a study are the same for every individual. We review these differences and the challenges they pose for ethical study design and implementation (Table 1).

Table 1. Challenges of research with diverse populations and potential solutions.

Ethical domains and challenges of diverse populations	Potential solutions
Recruiting representative populations	
• Different access to technology • Different social networks (eg, distinct networks of friends, community groups, and familial connections) • Size and cost of study	• Follow the principle of "Nothing about us without us" [39] • Form community advisory boards and engage with liaisons to help guide recruitment • Conduct recruitment alongside community participation and capacity building [40]
Informed consent	
• Different understandings of technology • Different disclosure practices • Translations, including American Sign Language, braille	• Multiple information sheets by language • Prior engagement toward the cocreation of knowledge • Differential practices of risk disclosure that balance informativeness and understanding • Targeted resources that enhance the consent process for vulnerable subpopulations
Secondary use of data	
• Heightened concerns for data reuse given historical exploitation and data misuse • Different expectations about the reuse of data, for example, by age	• Explicit notification and assent or withdrawal before deidentification for highly sensitive research • Community representation on decision-making bodies
Privacy	
• Increased data sensitivity (eg, disability, race, and gender identity) • High reidentification risk for hyperminoritized participants • Stigmatizing information that may be exacerbated for some subpopulations	• Separate demographic markers from primary subject data to prevent reidentification • Differential privacy techniques for stigmatizing research • Enhanced privacy and confidentiality practices
Risk to subjects	
• Specialized data risks to some subjects (eg, undocumented or pregnant persons) • Different cognitive capacities to appreciate or absorb risks	• Targeted research resources that differentially minimize the risk of participation • Inclusion of high-risk participants only as a last resort [41]
Risk to groups	
• Stigma for subpopulations generated by incidental research findings (eg, differential rates of alcohol or drug use)	• Separate demographic markers from primary subject data
Benefits to subjects	
• Different interests (ie, the value of devices and digital services) • Different social and technical capacities (ie, internet access and technical platform)	• Targeted research resources to equalize the benefits of participation • Explicit transfer of data collection technologies to individuals (ie, transfer and training in the use of tablets and laptops at the conclusion of the study)
Social value of research	
• Different interests and values • Different vulnerabilities to risks of research results (eg, group stigma and state oppression) • Limitation of benefits from research (eg, political or economic power)	• Explicit analysis of benefits by relevant subgroups • Explicit exploration, discussion, and co-created knowledge and transfer to policy for broad and just benefits

Differential Consent

Individuals may have very different expectations about digital tools and data and require different thresholds and procedures for participant consent. This may be a function of (1) understanding of the scope of participation, (2) tolerance for specific research purposes, and (3) the opportunity to decline participation if data collection occurs as part of routine clinical care or application usage. Each of these may influence the information that a participant requires, how researchers gather consent, and the kinds of research that can be undertaken with a waiver of consent.

Literacy about data collection and reuse is variable. Many individuals may not be aware that data about their physical location or browsing history is routinely collected [42]. Persons

with limited experience with digital technologies may have difficulty understanding the capacities, limits, and mechanisms underlying new technologies [43]. Moreover, expectations about the reuse of data may or may not shift depending on whether they are interacting with health care providers, nonprofit organizations, or private commercial entities (eg, Middleton et al [44] on attitudes toward genomic data sharing).

Individuals will have different standards for when risks should be disclosed or consent sought for specific research purposes. For example, in the case of contact tracing using digital apps that involve trade-offs between balancing public benefit and invasion of privacy [45], some individuals may be unwilling to have their deidentified data used if the research would inadvertently stigmatize their racial, gender, or religious group. Thus, while many people may accept deidentified data use for almost all research purposes, some may be particularly sensitive to undisclosed secondary research without their consent. Often marginalized or vulnerable groups may be especially concerned or cautious. Disagreement about research purposes will generate disagreement about what requires specific consent.

If consent to research is incorporated as part of nonresearch interactions (eg, clinical interactions or as part of the license agreement to use a digital service), then different individuals will have different capacities to decline. Access to health services is highly constrained for some groups, and some people may be uncomfortable declining research participation if consent is sought at the point of care. Likewise, awareness of the terms of user agreements to access digital devices and services is notoriously poor [46]. This unequal ability to refuse services complicates the assessment of which consent practices are justifiable.

Differential Risks and Benefits of Participation

Digital data have a high risk of stigmatizing subjects, placing their employment, income, and immigration status at risk, but these risks are not equal across all groups. In particular, individuals have different vulnerabilities to reidentification and to the harms of confidentiality breaches.

The risk for reidentification [47,48] is particularly high for individuals who are members of a hyperminority (ie, small ethnic subpopulations) or are members of multiple minority classes (ie, intersectionally marginalized). Certain kinds of individual behaviors such as widespread internet usage without privacy-preserving safeguards can also increase the risk of reidentification through commercially available databases, including for noncitizens, women, and youth [49].

Different individuals will have different vulnerability to harm resulting from reidentification. For example, people who are able to become pregnant are subject to special legal risks if reidentification exposes information about their reproductive health [50]. Likewise, individuals with marginalized sexualities or gender identities face differential risk of legal, social, and employment consequences from reidentification depending on their location, social context, and the prevailing legal environment [51-53]. These risks may be hard to assess or control in a deidentified data set that unlinks individuals from their data.

Differential Social Value

The social value of a study is dependent upon a person's capacity to access the benefits of the research. Different populations have different abilities to access the products of research, and a study may, therefore, produce insights into a particular condition that does not yield anticipated benefits. The result is a changed risk-benefit calculation for individual community members.

Maximizing social value also requires careful consideration of the different ways individuals assess the risks and benefits of the knowledge generated by research. For example, consider research on early detection of Alzheimer disease using data from mobile devices [54]. False negatives in diagnosis or screening can delay treatment, and false positives can generate unwarranted costs, anxiety, or stigma. Making a fine-grained trade-off between the sensitivity and specificity of a predictor is thus a decision about which people can reasonably disagree. However, it is often difficult to set individualized thresholds for false positives and false negatives. Maximizing social value for a predictor thus requires selecting an appropriate trade-off between adverse events, which differ by severity, frequency, and life impact across different subpopulations.

There are also different levels of vulnerability to the misuse of the research. For example, studies using digital tools (ie, social media data analysis) may generate incidental findings about the association between negatively valenced behaviors (such as suicidality, drug consumption, and negative moods) and marginalized racial groups, religions, sexualities, or genders. Moreover, the risk is asymmetric: the findings that run contrary to invidious stereotypes may not undermine popular beliefs in those stereotypes, even as findings, which support the stereotype reinforce the popular belief [55]. Marginalized communities suffer more from research that is potentially stigmatizing because they may start with fewer resources and opportunities.

Traditional Principles in the Context of Diversity

Traditional principles are workable in a context where the research population is relatively homogenous—where risks and benefits are similar or easy to calculate, where the information required to consent is easily accessible, and where values and mental models are uniform. With variability on any of the three dimensions, problems may appear:

1. Appropriately informing participants becomes more difficult as the vulnerabilities, languages, cultural understandings, and mental models of technology become more diverse. Diversity generates a dilemma for the information provision function of informed consent, by requiring investigators to either: (i) specify the full range of risks to which any member of the research population could reasonably be exposed or (ii) tailor information materials and procedures for each sufficiently distinct subgroup [56,57]. In the first strategy, the autonomy of subjects may be undermined by risk information that is irrelevant to their circumstances, potentially overloading the capacity of some individuals to meaningfully digest or understand risks. In the second

strategy, the burden upon researchers becomes more significant as the diversity of the group increases. More importantly, providing different information to subjects enrolled in the same study may violate respect for persons. Moreover, if researchers elide or simplify risks to a subject group because they believe the subjects are, on average, unlikely to be exposed to the risk, there is a serious risk that they will fail to fully inform members of that group [58].

2. Calculating the risk-benefit ratio for subjects becomes more difficult as the vulnerabilities, interests, and values of research participants become more diverse. According to the prevailing view in research ethics, some subjects should be excluded from studies, even if they would otherwise consent, if their participation would pose a sufficiently grave risk to their well-being. As the overlapping dimensions of diversity increase, so does the risk that at least some subjects are at high enough risk that they ought to be excluded [59]. Even innocuous data collection can place some members of a diverse research population at risk. For instance, the participation of undocumented persons [60], gender-diverse [52], or pregnant persons in seemingly innocuous research may expose location, financial, or health care data in ways that place them at risk of legal or social sanction in some jurisdictions. As diversity increases, however, it may become increasingly difficult to identify these individuals through simple demographic cues nor may participants feel comfortable reporting their status to researchers.

3. Calculating the social value of the research for the broader population requires considering the wide variety of different interests, religious and moral commitments, and material circumstances present within a plural society. Research impacts not only direct participants but also those who pay for, consume, benefit from, and live with the consequences of the research [15]. The social value requirement is sometimes understood as a minimal requirement that research has the potential to contribute valuable knowledge [61], while others view it as a more stringent requirement that research contributes to the maintenance of a just society [20]. At the very least, the social value of a research study cannot be thought of as a monolith: in a plural society, the costs and benefits of research will be different for different subpopulations [15]. If this is true, then the ethics of a research project must attend to the ways that research benefits and harms are distributed unequally. This is true not only for digital research with diverse populations but for all digital research.

Reorienting Toward Justice

The transactional model of researcher-participant ethics has created many of the challenges described so far: diverse research populations are addressed through individualized assessments that require individual consent and favorable individual risk-benefit. If any of these features fail, researchers are inclined to exclude individuals from a study. The rationales offered are:

1. Individualized consent is labor intensive, and subject diversity increases the risk that standardized consent practices will not adequately inform all participants. Researchers who wish to keep consent procedures manageable thus face a dilemma between, on the one hand, underdescribing research risks for some subjects and, on the other hand, overwhelming subjects with difficult-to-understand information.

2. While diversity raises the risk that at least some individuals will be harmed by participation, excluding those individuals because of their different risk profile may diminish the representativeness of data. Hence, there appears to be a trade-off between maximally protecting each individual subject and creating inclusive research projects that yield knowledge of benefit to a diverse society [62].

3. Some research designs, such as secondary reuse, do not permit individualized risk assessment or consent practices. Many risks arise in the context of deidentified secondary reuse, where (i) individuals can be reidentified, especially if in a minority population and (ii) where population-level findings can harm minoritized populations. Yet, these kinds of research do not allow for the individualized exclusion of subjects, and hence risks must be managed at the population level.

A complementary or alternative model is required. A reorientation of digital research ethics to emphasize social justice is supported by an overlapping consensus between recent work that explicitly considers the broader social and institutional ramifications of research. London [19] and Wenner [20] have argued that research is a system of social cooperation rather than a transaction between participant and researcher. Research should instead be understood as a collective activity that individuals engage in to protect and promote their basic interests through the shared production of a unique public good (ie, research knowledge). Likewise, the concept of data solidarity has been proposed by Prainsack et al [63] to emphasize collective control over research—private or public—to ensure that data are used in the public interest. This contrasts with the transactional model, which relies on a form of data altruism from subjects that treat digital data collection as a gift by subjects to researchers and fail to acknowledge the power imbalances between data subjects and data collectors, the harms that research can do to nonsubjects, and the potential for private profiteering from public data sets. Finally, participatory research methods reimagine participants as coresearchers, enabling the creation of new knowledge while meaningfully building capacity in affected communities [64]. These methods have been operationalized by indigenous and First Nations communities through the CARE framework and the OCAP principles, which focus on community control and capacity building beyond the direct subjects of the research.

These existing approaches adopt the tools of political philosophy to argue that research should not merely respect individual participants but also ensure a just allocation of the benefits and burdens of research within the community as a whole. While we cannot provide a full justification of these approaches in this paper, a way of specifying rights and obligations with respect to research is to consider the principles of research ethics to a community would agree to if they were unaware of their existing social roles, talents, wealth, or substantive moral commitments

[65]. This contractarian approach to justice has a venerable history among moral and political philosophers [65-67] and has been articulated in the context of health care justice [68] and research ethics [19,20]. On this view, the rules governing digital research would be structured so that the production of research knowledge protects the basic interests shared by all individuals in society, with special attention paid to the flourishing of the least advantaged members of society. This agreement would specify (1) the procedures to protect the autonomy of each individual to control participation in digital research, (2) general principles for fairly distributing the benefits and burdens of digital research, where deviations from a roughly equal distribution are permissible only if they are to the advantage of the most vulnerable, (3) a method for ensuring the legitimacy of the research and allowing for participants and affected community members to exercise control over crucial features of the research, and (4) that if the proceeding conditions are met, then there is a collective obligation to engage in the production of knowledge that serves to create a society that enables the flourishing of all people regardless of their advantages or disadvantages. While this does not yet specify concrete practices, it provides an alternative moral framework to the Belmont principles for justifying the development of guidelines for digital research with diverse populations.

Specific Recommendations

Overview

Drawing on this existing work, we outline considerations for good practices upon which researchers can draw and that will promote trust in the digital research process across diverse participant populations. While individual researchers cannot make the institutional and regulatory changes required to align with a justice-first view of digital research ethics, we believe that individual researchers can and should begin to attend to the special challenges of working with diverse study populations. The following considerations suggest that the digital research community should collectively work toward a fuller specification of justice that translates the existing work into the specific modalities of digital research with diverse populations.

Protect Individual Autonomy in High-Risk Situations

Digital research often exposes different subjects to different autonomy, privacy, and reputational risks. These are not well mitigated by standard collection and consent procedures, and the risks themselves are often invisible to researchers. As Prainsack et al [63] point out, as the risks to individuals increase or the public value of the research diminishes, collective control of data sets and individual rights to refuse should be strengthened. Researchers should thus consider the following practices:

1. Use a community advisory board and community liaisons throughout the study design process to identify high-risk activities and subpopulations requiring extra review and tailored consent processes.
2. Implement linguistically tailored consent processes for sufficient large or sufficiently vulnerable subpopulations. While this may increase the burden on researchers and oversight boards to vet additional consent documents, it is

an important method for showing respect for the participation of individuals who do not primarily communicate in the dominant language of the research population.
3. When engaging with people who hold divergent prior understandings of technology, prepare multiple risk disclosure strategies to balance informativeness and understanding and target study resources to enhance the consent process for more vulnerable subpopulations.
4. Explore institutional mechanisms and regulatory changes to allow for recontact to allow opportunities for withdrawal, including from secondary research with high risks for reidentification or community stigmatization. Given the privacy risks associated with maintaining data links to participants to enable recontact, the continued development of technical and sociotechnical practices for the preservation of participant privacy is essential [69].

Fairly Distribute Benefits and Burdens

Including a more diverse research population is an important method for a wide distribution of the benefits of research but may also burden already disadvantaged groups. While researchers cannot eliminate structural injustices completely, they should compensate for them in the design and conduct of their studies where possible. While there is likely to be disagreement over what methods of compensation are fair, a widely endorsed approach is to embrace methods that are to the benefit of the least advantaged members of society. This may include the following:

1. Target research resources to differentially minimize uneven risks of participation and barriers to research access. This may include capacity building among participating communities (eg, knowledge building around digital technologies).
2. In research with a high risk of reidentification or stigmatization, take care to enhance subject and community privacy. This may include differential privacy techniques [69] and the separation of demographic markers from primary subject data to prevent the reidentification of individuals and incidental stigmatization of marginalized populations.
3. Target research resources that equalize the benefits of participation for different groups with sensitivity to their different needs and capacities. Be explicit and proactive in developing mechanisms for disseminating results and policy mechanisms to ensure broad benefits of resulting knowledge.

Embrace Transparency and Build Collective Control of Research

Research often requires value judgments, and there may be widespread disagreement between researchers, participants, and the broader community over which values to promote. While resolving this disagreement is not simple, at the very least, it requires researchers to be transparent about their purposes, methods, and results. More controversially, but in alignment with the ethos of participatory research methods, it may require institutional mechanisms for community members to control

the direction of the research, invigilate risks, and ultimately steward data sets. This requires that researchers:

1. Clearly report research ethics methodologies, including specifying methods of consent, review of risks to participants, reporting of adverse events, and expected costs and benefits of producing the knowledge [3,70,71].

2. Incorporate community liaisons as members of the study team to identify connections with prospective participants and populations, tailoring consent documents, meaningful data analysis, and capacity building.

3. Educate and empower community advisory boards, especially for high-risk studies, to share in study design or development of study protocols.

4. Consider mechanisms for community ownership of data sets and strategies for limiting access to secondary reuse if deemed appropriate by community advisory boards or participants' representatives [12].

Implementation of all of these practices requires time, expertise, and money. The number of strategies that will need to be adopted is proportional to the expected balance of benefits and harms to participants and their communities [72]. The goal of doing research well, in ways that honor the trust that communities place in researchers, is non-negotiable.

Conclusions

The explicit integration of justice in digital research recognizes that societies have a collective duty to help produce valuable knowledge under fair terms of cooperation. While more work needs to be done to fully explicate the institutional changes and individual practices that will realize justice in digital research, existing work in participatory action research, neuroethics, political philosophy, and indigenous research methods already provide an invaluable foundation for discharging this obligation.

Conflicts of Interest

None declared.

References

1. Eaneff S, Obermeyer Z, Butte A. The case for algorithmic stewardship for artificial intelligence and machine learning technologies. JAMA 2020;324(14):1397-1398 [doi: 10.1001/jama.2020.9371] [Medline: 32926087]

2. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science 2019;366(6464):447-453 [FREE Full text] [doi: 10.1126/science.aax2342] [Medline: 31649194]

3. Sounderajah V, McCradden MD, Liu X, Rose S, Ashrafian H, Collins GS, et al. Ethics methods are required as part of reporting guidelines for artificial intelligence in healthcare. Nat Mach Intell 2022;4(4):316-317 [doi: 10.1038/s42256-022-00479-3]

4. Tindale LC, Chiu D, Minielly N, Hrincu V, Talhouk A, Illes J. Wearable biosensors in the workplace: perceptions and perspectives. Front Digit Health 2022;4:800367 [FREE Full text] [doi: 10.3389/fdgth.2022.800367] [Medline: 35860688]

5. Claw KG, Anderson MZ, Begay RL, Tsosie KS, Fox K, Garrison NA, Summer internship for INdigenous peoples in Genomics (SING) Consortium. A framework for enhancing ethical genomic research with Indigenous communities. Nat Commun 2018;9(1):2957 [FREE Full text] [doi: 10.1038/s41467-018-05188-3] [Medline: 30054469]

6. Harding L, Marra CJ, Illes J. Establishing a comprehensive search strategy for Indigenous health literature reviews. Syst Rev 2021;10(1):115 [FREE Full text] [doi: 10.1186/s13643-021-01664-y] [Medline: 33874997]

7. Harding L, Marra CJ, Manohara V, Illes J. Ways of knowing of the brain and mind: a scoping review of the literature about global indigenous perspectives. J Neurol Res 2022 Aug;12(2):43-53 [doi: 10.14740/jnr708]

8. Harding L, Illes J. RE: Canadian assessment of deep brain stimulation access: the Canada study. Can J Neurol Sci 2021;48(1):130-131 [FREE Full text] [doi: 10.1017/cjn.2020.151] [Medline: 32669146]

9. Harding L, McFarlane J, Honey CR, McDonald J, Illes J. Mapping the landscape of equitable access to advanced neurotechnologies in Canada. Can J Neurol Sci 2023;50(s1):s17-s25 [doi: 10.1017/cjn.2023.18] [Medline: 37160675]

10. Perreault ML, King M, Gabel C, Mushquash CJ, De Koninck Y, Lawson A, et al. An Indigenous lens on priorities for the Canadian brain research strategy. Can J Neurol Sci 2023;50(1):96-98 [FREE Full text] [doi: 10.1017/cjn.2021.501] [Medline: 34847973]

11. Research Data Alliance International Indigenous Data Sovereignty Interest Group. CARE Principles for Indigenous Data Governance. The Global Indigenous Data Alliance. 2019. URL: https://GIDA-global.org [accessed 2023-08-16]

12. The First Nations Information Governance Centre. Ownership, Control, Access and Possession (OCAP: the Path to First Nations Information Governance (Paper). Ottawa: The First Nations Information Governance Centre; 2014. URL: https://www.deslibris.ca/ID/10095457 [accessed 2022-08-04]

13. MacLeod H, Jacobs M, Siek K, Connelly K, Mynatt ED. Ethical Considerations in Pervasive Health Research. Mexico: ACM; 2016 Presented at: 10th EAI International Conference on Pervasive Computing Technologies for Healthcare; June 16, 2016; Mexico URL: https://eudl.eu/doi/10.4108/eai.16-5-2016.2263813 [doi: 10.4108/eai.16-5-2016.2263813]

14. National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. The Belmont Report. Department of Health, Education and Welfare. 1979. URL: https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/read-the-belmont-report/index.html [accessed 2021-03-26]

15. Herington J, Tanona S. The social risks of science. Hastings Cent Rep 2020;50(6):27-38 [doi: 10.1002/hast.1196] [Medline: 33315249]

16. MacKay D. Fair subject selection in clinical and social scientific research. In: Iltis AS, MacKay D, editors. The Oxford Handbook of Research Ethics. Oxford: Oxford University Press; 2020.

17. Pratt B, Wild V, Barasa E, Kamuya D, Gilson L, Hendl T, et al. Justice: a key consideration in health policy and systems research ethics. BMJ Glob Health BMJ Spec J 2020;5(4):e001942 [FREE Full text] [doi: 10.1136/bmjgh-2019-001942]

18. Loignon C, Dupéré S, Leblanc C, Truchon K, Bouchard A, Arsenault J, et al. Equity and inclusivity in research: co-creation of a digital platform with representatives of marginalized populations to enhance the involvement in research of people with limited literacy skills. Res Involv Engagem 2021;7(1):70 [FREE Full text] [doi: 10.1186/s40900-021-00313-x] [Medline: 34610847]

19. London AJ. For the Common Good: Philosophical Foundations of Research Ethics. Oxford, New York: Oxford University Press; 2021.

20. Wenner DM. The social value requirement in research: from the transactional to the basic structure model of stakeholder obligations. Hastings Cent Rep 2018;48(6):25-32 [doi: 10.1002/hast.934] [Medline: 30586181]

21. Angwin J, Larson J. Bias in criminal risk scores is mathematically? ProPublica. 2016. URL: https://www.propublica.org/article/bias-in-criminal-risk-scores-is-mathematically-inevitable-researchers-say [accessed 2017-11-02]

22. Buolamwini J, Gebru T. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. 2018 Presented at: Proceedings of the 1st Conference on Fairness, Accountability and Transparency; 2018; Cambridge, MA, New York, NY p. 77-91 URL: http://proceedings.mlr.press/v81/buolamwini18a.html

23. McFarlane J, Illes J. Neuroethics at the interface of machine learning and schizophrenia. NPJ Schizophr 2020;6(1):18 [FREE Full text] [doi: 10.1038/s41537-020-0108-6] [Medline: 32686681]

24. Bissoto A, Fornaciali M, Valle E, Avila S. (De)constructing bias on skin lesion datasets. : IEEE; 2019 Presented at: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW); June 16-17, 2019; Long Beach, CA, USA p. 2766-2774 URL: https://ieeexplore.ieee.org/document/9025695/authors#authors [doi: 10.1109/cvprw.2019.00335]

25. Hellman D. Measuring algorithmic fairness. V Law Rev 2020;106(4):811-866

26. Colvonen PJ, DeYoung PN, Bosompra NOA, Owens RL. Limiting racial disparities and bias for wearable devices in health science research. Sleep 2020;43(10):zsaa159 [FREE Full text] [doi: 10.1093/sleep/zsaa159] [Medline: 32893865]

27. Guo LN, Lee MS, Kassamali B, Mita C, Nambudiri VE. Bias in, bias out: underreporting and underrepresentation of diverse skin types in machine learning research for skin cancer detection-a scoping review. J Am Acad Dermatol 2022;87(1):157-159 [doi: 10.1016/j.jaad.2021.06.884] [Medline: 34252465]

28. Lee K, Sener IN. Strava Metro data for bicycle monitoring: a literature review. Transp Rev Routledge 2021;41(1):27-47 [doi: 10.1080/01441647.2020.1798558]

29. Adepoju OE, Chae M, Ojinnaka CO, Shetty S, Angelocci T. Utilization gaps during the COVID-19 pandemic: racial and ethnic disparities in telemedicine uptake in federally qualified health center clinics. J Gen Intern Med 2022;37(5):1191-1197 [FREE Full text] [doi: 10.1007/s11606-021-07304-4] [Medline: 35112280]

30. Chunara R, Zhao Y, Chen J, Lawrence K, Testa PA, Nov O, et al. Telemedicine and healthcare disparities: a cohort study in a large healthcare system in New York City during COVID-19. J Am Med Inform Assoc 2021;28(1):33-41 [FREE Full text] [doi: 10.1093/jamia/ocaa217] [Medline: 32866264]

31. Lewis JM. Internet access and racial/ethnic disparities in using internet health resources. Report No.: SESHD 2017-31. US Census Bureau. 2017. URL: https://www.census.gov/library/working-papers/2017/demo/SEHSD-WP2017-31.html [accessed 2023-08-16]

32. Pew Research Center. Pew Res Cent Internet Sci Tech. Demographics of Internet and Home Broadband Usage in the United States. 2021. URL: https://www.pewresearch.org/internet/fact-sheet/internet-broadband/ [accessed 2022-06-09]

33. Zahnd WE, Bell N, Larson AE. Geographic, racial/ethnic, and socioeconomic inequities in broadband access. J Rural Health 2022;38(3):519-526 [doi: 10.1111/jrh.12635] [Medline: 34792815]

34. Makri A. Bridging the digital divide in health care. Lancet Digit Health 2019;1(5):e204-e205 [FREE Full text] [doi: 10.1016/s2589-7500(19)30111-6]

35. Ortega G, Rodriguez JA, Maurer LR, Witt EE, Perez N, Reich A, et al. Telemedicine, COVID-19, and disparities: policy implications. Health Policy Technol 2020;9(3):368-371 [FREE Full text] [doi: 10.1016/j.hlpt.2020.08.001] [Medline: 32837888]

36. Protection of Human Subjects (HHS Common Rule). CFR. Sect Part 46-Protection of Human Subjects Jan 19. 2017. URL: https://www.ecfr.gov/cgi-bin/retrieveECFR?gp=&SID=83cd09e1c0f5c6937cd9d7513160fc3f&pitd=20180719&n=pt45. 1.46&r=PART&ty=HTML [accessed 2018-09-12]

37. Him DA, Aguilar TA, Frederick A, Larsen H, Seiber M, Angal J. Tribal IRBs: a framework for understanding research oversight in American Indian and Alaska Native Communities. Am Indian Alsk Native Ment Health Res 2019;26(2):71-95 [FREE Full text] [doi: 10.5820/aian.2602.2019.71] [Medline: 31550379]

38. Canadian Institutes of Health Research, Natural Sciences and Engineering Research Council of Canada, Social Sciences and Humanities Research Council of Canada. Tri-council policy statement: ethical conduct for research involving humans. Ottawa, Canada: Secretariat on Responsible Conduct of Research; 2018. URL: https://ethics.gc.ca/eng/documents/ tcps2-2018-en-interactive-final.pdf [accessed 2023-08-16]

39. Charlton JI. Nothing About Us Without Us: Disability Oppression and Empowerment. Oakland, CA: University of California Press; 2000.

40. Stevenson S, Bruce L, Dwosh E, Beattie BL, Illes J. When culture informs neuroscience: considerations for community-based neurogenetics research and clinical care in a first nation community with early onset familial alzheimer disease. In: Chiao JY, editor. The Oxford Handbook of Cultural Neuroscience. Oxford: Oxford Library of Psychology; 2015:171-182

41. Eijkholt M, Anderson JA, Illes J. Picturing neuroscience research through a human rights lens: imaging first-episode schizophrenic treatment-naive individuals. Int J Law Psychiatry 2012;35(2):146-152 [FREE Full text] [doi: 10.1016/j.ijlp.2011.12.003] [Medline: 22304987]

42. Valentino-DeVries J, Singer N, Keller MH, Krolik A. Your apps know where you were last night, and they're not keeping it secret. N Y Times. 2018. URL: https://www.nytimes.com/interactive/2018/12/10/business/location-data-privacy-apps. html [accessed 2022-08-03]

43. Chiu CJ, Liu CW. Understanding older adult's technology adoption and withdrawal for elderly care and education: mixed method analysis from national survey. J Med Internet Res 2017;19(11):e374 [FREE Full text] [doi: 10.2196/jmir.7401] [Medline: 29101093]

44. Middleton A, Milne R, Almarri MA, Anwer S, Atutornu J, Baranova EE, et al. Global public perceptions of genomic data sharing: what shapes the willingness to donate DNA and health data? Am J Hum Genet 2020;107(4):743-752 [FREE Full text] [doi: 10.1016/j.ajhg.2020.08.023] [Medline: 32946764]

45. Gasser U, Ienca M, Scheibner J, Sleigh J, Vayena E. Digital tools against COVID-19: taxonomy, ethical challenges, and navigation aid. Lancet Digit Health 2020;2(8):e425-e434 [FREE Full text] [doi: 10.1016/S2589-7500(20)30137-0] [Medline: 32835200]

46. Litman-Navarro K. Opinion: We Read 150 Privacy Policies: They Were an Incomprehensible Disaster. N Y Times. 2019. URL: https://www.nytimes.com/interactive/2019/06/12/opinion/facebook-google-privacy-policies.html [accessed 2022-08-09]

47. 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]

48. Xia W, Liu Y, Wan Z, Vorobeychik Y, Kantacioglu M, Nyemba S, et al. Enabling realistic health data re-identification risk assessment through adversarial modeling. J Am Med Inform Assoc 2021;28(4):744-752 [FREE Full text] [doi: 10.1093/jamia/ocaa327] [Medline: 33448306]

49. Burkell J, Regan P. Expression in the virtual public: social justice considerations in harvesting youth online discussions for research purposes. SSJ 2021;15(3):397-413 [doi: 10.26522/ssj.v15i3.2536]

50. Torchinsky R. How period tracking apps and data privacy fit into a post-Roe v. Wade climate. 2022. URL: https://www. npr.org/2022/05/10/1097482967/roe-v-wade-supreme-court-abortion-period-apps [accessed 2022-06-13]

51. Alpert AB, Komatsoulis GA, Meersman SC, Garrett-Mayer E, Bruinooge SS, Miller RS, et al. Identification of transgender people with cancer in electronic health records: recommendations based on CancerLinQ observations. JCO Oncol Pract 2021;17(3):e336-e342 Wolters Kluwer [FREE Full text] [doi: 10.1200/OP.20.00634] [Medline: 33705680]

52. Kronk CA, Everhart AR, Ashley F, Thompson HM, Schall TE, Goetz TG, et al. Transgender data collection in the electronic health record: current concepts and issues. J Am Med Inform Assoc 2022;29(2):271-284 [FREE Full text] [doi: 10.1093/jamia/ocab136] [Medline: 34486655]

53. Lett E, Everhart A. Considerations for transgender population health research based on US national surveys. Ann Epidemiol 2022;65:65-71 [doi: 10.1016/j.annepidem.2021.10.009] [Medline: 34757013]

54. Frey AL, Karran M, Jimenez RC, Baxter J, Adeogun M, Bose N, et al. Harnessing the potential of digital technologies for the Early Detection of Neurodegenerative Diseases (EDoN). OSF Preprints 2019:1-15 [FREE Full text] [doi: 10.31219/osf.io/u49z5]

55. Kitcher P. Science, Truth, and Democracy. Revised edition. Oxford: Oxford University Press; 2003.

56. Bull S, Farsides B, Ayele FT. Tailoring information provision and consent processes to research contexts: the value of rapid assessments. J Empir Res Hum Res Ethics 2012;7(1):37-52 [FREE Full text] [doi: 10.1525/jer.2012.7.1.37] [Medline: 22378133]

57.   Tekola F, Bull SJ, Farsides B, Newport MJ, Adeyemo A, Rotimi CN, et al. Tailoring consent to context: designing an appropriate consent process for a biomedical study in a low income setting. PLoS Negl Trop Dis 2009;3(7):e482 [FREE Full text] [doi: 10.1371/journal.pntd.0000482] [Medline: 19621067]

58.   Bromwich D, Rid A. Can informed consent to research be adapted to risk? J Med Ethics 2015;41(7):521-528 [FREE Full text] [doi: 10.1136/medethics-2013-101912] [Medline: 25351375]

59.   Ries NM, Thomson M. Bioethics and universal vulnerability: exploring the ethics and practices of research participation. Med Law Rev 2020;28(2):293-316 [FREE Full text] [doi: 10.1093/medlaw/fwz026] [Medline: 32524142]

60.   Jach E, Gloeckner G, Kohashi C. Social and behavioral research with undocumented immigrants: navigating an IRB committee. Hispanic J Behav Sci 2020;42(1):3-17 [FREE Full text] [doi: 10.1177/0739986319899979]

61.   Wendler D, Rid A. In defense of a social value requirement for clinical research. Bioethics 2017;31(2):77-86 [FREE Full text] [doi: 10.1111/bioe.12325] [Medline: 28060427]

62.   Friesen P, Gelinas L, Kirby A, Strauss DH, Bierer BE. IRBs and the protection-inclusion dilemma: finding a balance. Am J Bioeth 2023;23(6):75-88 [doi: 10.1080/15265161.2022.2063434] [Medline: 35482887]

63.   Prainsack B, El-Sayed S, Forgó N, Szoszkiewicz L, Baumer P. Data solidarity: a blueprint for governing health futures. Lancet Digit Health 2022;4(11):e773-e774 Elsevier [FREE Full text] [doi: 10.1016/S2589-7500(22)00189-3] [Medline: 36307191]

64.   Lenette C. Participatory Action Research: Ethics and Decolonization. New York, NY: Oxford University Press; 2022.

65.   Rawls J. A Theory of Justice. Rev. Cambridge: Belknap Press; 1999.

66.   Pateman C, Mills CW. In: Mills CW, editor. Contract and Domination. Cambridge: Polity; 2007.

67.   Scanlon TM. What We Owe to Each Other. Cambridge: Belknap Press; 1998.

68.   Daniels N. Just Health: Meeting Health Needs Fairly. New York: Cambridge University Press; 2008.

69.   Fast-Track Action Committee on Advancing Privacy-Preserving Data Sharing and Analytics. National strategy to advance privacy-preserving data sharing and analytics. Washington, D.C: National Science and Technology Council; 2023. URL: https://www.whitehouse.gov/wp-content/uploads/2023/03/National-Strategy-to-Advance-Privacy-Preserving-Data-Sharing-and-Analytics.pdf [accessed 2023-07-07]

70.   Dwork C, Roth A. The algorithmic foundations of differential privacy. Found Trends® Theor Comput Sci 2014;9(3-4):211-407 Now Publishers, Inc. [doi: 10.1561/0400000042]

71.   Anderson JA, Eijkholt M, Illes J. Ethical reproducibility: towards transparent reporting in biomedical research. Nat Methods Nat 2013;10(9):843-845 Publishing Group. [doi: 10.1038/nmeth.2564] [Medline: 23985730]

72.   Beauvais MJS, Knoppers BM, Illes J. A marathon, not a sprint—neuroimaging, open science and ethics. Neuroimage 2021;236:118041 [FREE Full text] [doi: 10.1016/j.neuroimage.2021.118041] [Medline: 33848622]

Abbreviations

CARE: Collective benefit, Authority to Control, Responsibility, and Ethics
OCAP: ownership, control, access, and possession

XSL•FO
RenderX