

UC Berkeley

UC Berkeley Previously Published Works

Title

Advancing digital sensing in mental health research.

Permalink

<https://escholarship.org/uc/item/5tb919ds>

Journal

npj Digital Medicine, 7(1)

Authors

Akre, Samir

Seok, Darsol

Douglas, Christopher

et al.

Publication Date

2024-12-18

DOI

10.1038/s41746-024-01343-x

Peer reviewed

<https://doi.org/10.1038/s41746-024-01343-x>

Advancing digital sensing in mental health research

Check for updates

Samir Akre^{1,2}, Darsol Seok^{2,3}, Christopher Douglas^{2,4}, Adrian Aguilera^{5,6}, Simona Carini⁷, Jessilyn Dunn⁸, Matthew Hotopf⁹, David C. Mohr¹⁰, Alex A. T. Bui^{1,2,11} & Nelson B. Freimer^{2,12} ✉

Digital sensing tools, like smartphones and wearables, offer transformative potential for mental health research by enabling scalable, longitudinal data collection. Realizing this promise requires overcoming significant challenges including limited data standards, underpowered studies, and a disconnect between research aims and community needs. This report, based on the 2023 Workshop on Advancing Digital Sensing Tools for Mental Health, articulates strategies to address these challenges to ensure rigorous, equitable, and impactful research.

The data obtainable from the sensors on digital devices—including smartphones, smart watches, and other wearables – have the potential to transform research on mental health disorders by informing our understanding of their causes and enabling us to delineate underlying subtypes, predict their diverse trajectories, and identify potential points of intervention through observational studies and clinical trials. The field of mental health research, from its beginnings in the 19th Century, has relied on subjective assessments, principally from self-report of symptoms and experiences, both to classify these disorders and to assess their course; these classifications are imprecise and encompass considerable heterogeneity^{1–3}. More recently, researchers have had access to technologies such as neuroimaging and electroencephalography, which enable objective assessment of the relationships between emotions and behaviors and both physiologic and anatomic features. However, as these approaches have, to date, largely been limited to laboratory settings with limited ecological validity and have been difficult to scale, they have not yet demonstrated utility in providing more precise and less heterogeneous phenotypes for population-level mental health investigations.

Because consumer digital devices are now used pervasively and provide readouts on a wide range of behavioral and physiological variables, such as sleep, activity, and cardiovascular parameters like heart rate and heart rate variability, many researchers are optimistic that they will provide the first tools for objectively obtaining phenotypic assessments of mental health disorders at scale⁴. Such utility has not yet, however, been conclusively demonstrated. In this Perspective, we highlight key objectives that the field of digital mental health research must meet to fulfill its promise. We suggest

here strategies for doing so, drawing on the recommendations of a 2023 international workshop, “Advancing the Utility of Digital Sensing Tools for Mental Health Research.”

Because mental health disorders are enormously heterogeneous, studies aiming to elucidate their etiology, define subgroups relevant to treatment choices, or predict their course require very large samples as effect sizes are inherently low. To accomplish any of these aims, digital sensing in mental health must decisively move beyond the exploratory studies that still constitute most of the literature in this field⁵. Well-powered discovery and replication studies will require researchers to increase sample sizes by orders of magnitude compared to current practice.

One way to increase scale will be to combine data across studies, leveraging the increasingly global ubiquity of consumer digital devices. However such a scale up will require concerted efforts to overcome several factors that act as impediments to data combination efforts, including the following: First, device manufacturers typically modify both hardware and software at frequent intervals and currently do so in ways that are largely opaque to investigators⁶. Further, as variables may be measured or reported in different ways, researchers must find ways to combine heterogeneous measurements. Additionally, researchers may be reluctant to share digital device data given uncertainties regarding the risks of such sharing⁷. And finally, effective data combination efforts will likely also depend on our obtaining a greater understanding of how socioeconomic variables and other group differences may influence the relationship between aspects of device usage and psychological or mental health measures⁸.

¹Medical Informatics Home Area, University of California Los Angeles, Los Angeles, CA, USA. ²Depression Grand Challenge, University of California Los Angeles, Los Angeles, CA, USA. ³Interdepartmental Program in Neuroscience, University of California Los Angeles, Los Angeles, CA, USA. ⁴Department of Psychiatry, School of Medicine, Washington University in St Louis, St Louis, MO, USA. ⁵School of Social Welfare, University of California Berkeley, Berkeley, CA, USA. ⁶Department of Psychiatry and Behavioral Sciences, Zuckerberg San Francisco General Hospital, University of California San Francisco, San Francisco, CA, USA. ⁷Department of Medicine, University of California San Francisco, San Francisco, CA, USA. ⁸Biomedical Engineering Department, Duke University, Durham, NC, USA. ⁹Department of Psychological Medicine, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, United Kingdom. ¹⁰Center for Behavioral Intervention Technologies, Department of Preventive Medicine, Northwestern University, Chicago, IL, USA. ¹¹Medical & Imaging Informatics Group, University of California Los Angeles, Los Angeles, CA, USA. ¹²Center for Neurobehavioral Genetics, Semel Institute for Neuroscience and Human Behavior, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA. ✉e-mail: NFreimer@mednet.ucla.edu

Overall, however, we consider the failure of the field to achieve agreed-upon standards for generating, interpreting, and sharing data to have played a crucial role in limiting its progress and believe that we can learn valuable lessons from other life science fields, such as human genomics, that have overcome similar challenges, as described below.

The release of the first draft of the human genome sequence in 2001 led quickly to coordinated efforts by the genomics field, supported by the US National Institutes of Health (NIH) and other funders to identify millions of genetic variants; the principal purpose of these efforts was to enable the discovery of genetic associations with common diseases, one of the central objectives of the Human Genome Project. While the availability of these variants led, over the ensuing few years, to a rapid growth in the number of genetic association studies performed for both common diseases and quantitative traits, these studies initially generated only a handful of replicated associations. Over this time, it became apparent that the lack of recognized standards for such studies was limiting the progress of the field, as noted in several methodological publications. By 2006, this awareness inspired two institutes at the US National Institutes of Health (the National Human Genome Research Institute [NHGRI] and the National Cancer Institute [NCI]) to jointly assemble a working group, including researchers, journal editors, and NHGRI/NCI staff, to propose guidelines that could constitute accepted standards for the design and reporting of genetic associations⁹. As a result of these and other efforts, the field quickly adopted rigorous criteria for considering associations statistically significant (both initial reports and replication efforts). These criteria, in turn, stimulated the design of studies that would be adequately powered to achieve such criteria, which required the field to rapidly develop a culture of extensive data sharing, effected both through the formation of international research consortia¹⁰ and the development of easily accessible data repositories¹¹. Additionally, the need to analyze data across multiple studies hastened the spread of genotyping platforms that could inexpensively produce standardized data and the development and rapid acceptance of uniform analysis methods.

Taken together, the different steps that were followed to achieve standards for genetic association studies enabled a spectacular acceleration of progress, such that there are now millions of replicated associations for thousands of traits. These discoveries have had a tremendous impact across the biomedical sciences, from identification of biological mechanisms underlying diseases to clinical implementations for predictions of disease risk at the individual level¹².

2023 Workshop on Advancing the Utility of Digital Sensing Tools for Mental Health Research

Taking a leaf from the experience of human genomics, thought leaders from multiple disciplines (drawn from academia, industry, funding and regulatory agencies, and groups representing individuals living with mental health disorders) have begun to grapple with the task of ensuring that mental health research using digital devices is scalable, yields reproducible findings, and can be conducted in ways that are both acceptable to diverse populations of research participants and useful to the communities from which they are drawn. Achieving this objective requires first delineating obstacles that impede progress and then identifying and implementing practical steps needed for the field to overcome these obstacles. The 53 participants in the 2023 digital sensing workshop, which was sponsored by the US National Institute of Mental Health (NIMH) and Wellcome, addressed both steps.

The workshop considered the obstacles to sensor-based assessments across three broad levels (Digital Infrastructure & Data Flow; Study Design & Reporting; and User Perspectives, Fig. 1), and its participants met in different workgroups that proposed solutions specific to each of these levels and produced reports that were made available for public comment (https://escholarship.org/uc/ucla_depression_grandchallenge). Below, we summarize the main challenges and recommendations identified by each group.

Digital infrastructure and data flow

These two workgroups addressed how to effectively organize the wide range of sensors, apps, and data being used globally and how to harmonize

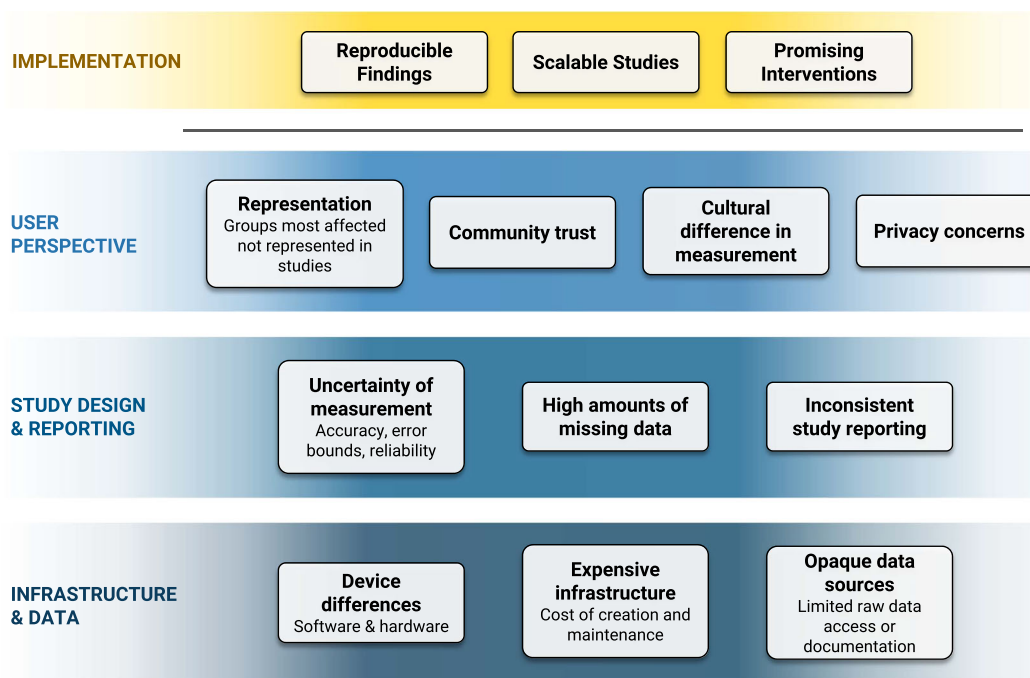


Fig. 1 | Overview of key issues discussed in the digital sensing workshop’s workgroups as they relate to key implementation goals for the field of behavioral digital sensing. The organizational structure shown here was adapted from a multilevel framework previously proposed by Mohr et al. for deriving and analyzing sensor features for mental health research⁴³. The lowest level considers the sensors themselves (e.g., gyroscopes or accelerometers), on top of which are the low-level

features derived from the sensors (e.g., types of activity or degrees of movement intensity), then higher-level behavioral markers (e.g., psychomotor activity or fatigue), and finally, clinical states (e.g., depression or anxiety). Figure 1 reframes the original figure in terms of key barriers to reaching goals in the field of behavioral digital sensing.

Box 1 | Digital infrastructure and data flow recommendations

1. Build standards for data from digital sensing devices
 - a. Address version changes and between-device differences for the same measure for hardware and software
 - b. Establish minimum requirements for a given measure (error bounds, sunset period, required metadata)
2. Fund Centers of Excellence centered on digital sensing for mental health to:
 - a. Create benchmark datasets
 - b. Maintain a network of sites with interdisciplinary expertise
3. Identify the common ground between device manufacturers, academics, clinicians, and users
 - a. Define highest potential impact areas for sensor development
 - b. Establish protocols for ensuring privacy of participant data
- c. Develop reusable assets for digital mental health studies starting with common APIs and endpoints

pipelines and methods for collecting data from devices, analyzing results, and sharing key findings (Box 1).

Device differences

Devices can be grouped by their intended use (clinical/research or consumer/wellness), although the lines between these uses are becoming more blurred every day^{6,13,14}. The differences between devices, not only across categories but even different versions of the same device, have impeded efforts to replicate the results of digital behavioral sensing studies. As these differences are usually not transparent to most researchers, progress in the field depends on investigators gaining greater access to information on sensor components and specifications as well as the variables that are reported by the device. Differences between consumer platforms may be particularly important barriers to the equitable implementation of digital behavioral sensing, as the relative frequency of use of specific platforms varies considerably according to geographic and sociodemographic factors.

The building of standards for the types, formats, and parameters to adequately characterize data from consumer devices is one step toward being able to meaningfully compare data within and across studies. This process should leverage existing efforts, such as the IEEE Standard 1752.1-2021¹⁵, IEEE Working Group on digital health (P1752 Open Mobile Health)¹⁶, DiMe's Digital Health Measurement Collaborative Community (DATAcc)¹⁷, and Open mHealth^{18,19}. To facilitate further community and collaboration, it may be valuable to develop a network of expertise and an online community accessible to those conducting research on digital sensing in mental health (along the lines of programs such as Biostars for the bioinformatics community²⁰) and to build on communities formed through the National Science Foundation (NSF) Smart and Connected Communities program²¹.

Expensive infrastructure

Specifying standards is only a first step. The field will need accessible infrastructure to enable implementation of such standards in the process of large-scale data generation. Some current studies are collecting data on the scale of terabytes per participant over long time periods (in some cases several years), and the costs and complexity of this infrastructure are thus increasing rapidly. Efforts on this scale are not supportable with infrastructure that has largely developed on a bespoke, per-project basis, given rapid changes in technologies, application programming interfaces (APIs), and underlying data/communication standards. For this reason, scalable deployment of digital sensing in mental health research is now limited to a few sites with outsized resources and expertise. Even for such sites, the current landscape for funding does not adequately support the long-term maintenance of infrastructure or distribution of assets critical for research.

One mechanism for obtaining a more sustainable and equitable distribution of necessary infrastructure would be the formation of collaborative Centers of Excellence focused on digital sensing for mental

health; establishing, supporting, and disseminating best practices in a sustainable manner; and curating shared datasets available to the scientific community. This strategy would incentivize groups to develop and maintain a set of reusable assets while pushing forward standards in the field. Key initiatives funded by this approach would be: 1) creating benchmark datasets; 2) maintaining a network of sites with interdisciplinary expertise; and 3) "priming the pump" for research through training and infrastructure buildout.

The creation of large benchmark datasets for digital sensing in mental health would simultaneously establish best practices that prioritize data privacy and sensitivity, while facilitating their dissemination to the research community. While mental health digital sensing data present distinct privacy concerns, other fields, such as genomics, have established standard practices for widespread sharing of similarly sensitive data. As device manufacturers may play a role in the generation of benchmark datasets, it will be important to engage these companies, from the outset of studies, in developing data sharing plans.

The interdisciplinary infrastructure of Centers of Excellence should initially focus on items that will remain scalable, such as common APIs and endpoints, and extend to include software for data aggregation and multimodal visualization (e.g., the Digital Biomarker Discovery Project²²). Additionally, relying on postdoctoral scholars and graduate students for the maintenance of software infrastructure on this scale is not feasible. It will therefore be necessary for Centers of Excellence to include funding for research engineers over multiyear timespans. Importantly, this framework should reward investigators who enable reproducible research, prioritize underserved and underrepresented groups, and implement FAIR²³ (Findable, Accessible, Interoperable, and Reusable) and TRUST²⁴ (Transparency, Responsibility, User focus, Sustainability and Technology) principles.

Opaque data sources

Device developers must navigate between their need for financially sustainable business models (e.g., proprietary and marketable technologies) and the demands from the research community for transparency in how their products function. The workgroup participants were emphatic in stating the view that fostering reproducible science does not require developers to reveal all their proprietary technology; indeed, it could strengthen their claims regarding the validity and efficacy of their devices. The participants further suggested that facilitating more extensive formal collaborations, throughout the process of developing and testing hardware and apps, could help bridge the gap between industry and academic sponsors. One of the workshop sponsors, the NIMH, has accepted this suggestion; in a recent announcement of a new funding opportunity, "Standardizing Data and Metadata from Wearable Devices," <https://grants.nih.gov/grants/guide/pa-files/PAR-24-250.html>, it strongly encouraged the inclusion of device manufacturers in research teams responding to this announcement.

Box 2 | Recommendations for digital sensing study design and reporting

Study design and participant recruitment

- Individuals with lived experience of mental health disorders should participate in the design of the study.
- Studies analyzing digital sensing data in relation to self-report instruments should collect information that will enable researchers to know when and in what situations participants are completing these instruments.
- Consider how contact between study staff and participants may impact study outcomes.
- Prioritize collection of data elements that can provide surrogate measures of data completeness or participant compliance with the study protocol. (e.g., hours of watch wear, phone off time)
- Mental health studies often include risk mitigation strategies for information gained from interviews or self-report assessments. Studies should consider what strategies may be warranted to mitigate the risk detected from digital sensing assessments.

Study information

Digital sensing studies should report:

- Setting, locations, and relevant dates, including periods of recruitment and data collection
- Processes used to promote participant adherence to the study protocol, including participant training, adherence monitoring, outreach processes, and participant compensation
- The sensor technology used, including make, model, versions, form factor, wear location, and sensor modalities (e.g., type, units, sampling rates, etc.)

- Missing data for all data types and techniques used to account for them
- Item-level responses in data from self-report instruments

Researchers should deposit data in access-controlled repositories such as PhysioNet or Sage Bionetworks.

Data quality issues

- Researchers should, if feasible, provide participants with the wearable devices from which study data are obtained as a means of controlling for device and sensor differences in software or hardware.
- The smartphones included in a study must be capable of collecting the sensor data of interest; therefore, researchers may consider the need to provide suitable phones and data plans to those who do not have them.
- Leverage existing digital infrastructure (data collection apps, data storage mechanisms) to avoid reinventing the wheel
- Routinely and continuously monitor data quality, including both missing data and outliers, to identify device and software failures or participants who are having difficulty adhering to protocols to allow for timely resolution of problems.

Results

- Assess algorithm and model performance stratified by demographic and clinical characteristics.
- Report limitations on how representative the dataset is to trained model generalizability

To reach a state where we understand how specific device hardware and software versions influence measurements of interest (and what those measurements are) will require channels of communication between digital device companies, researchers, clinicians, funders, and users with lived experience. The continuation of workshops such as the one that generated this report and the leveraging of existing spaces where funding agencies, industry, and academic groups may already meet can foster such engagement. These discussions can provide three main “products” that will facilitate scalable research: 1) information on how software and hardware versioning and updates influence the sensor data that researchers gather; 2) well-documented APIs for gathering digital sensing data; and 3) creation of standards for what comprises data fit for research²⁵. Additionally, as current consumer devices may not include some sensors that could have the greatest potential for impacting our understanding of mental health, efforts to foster iterative discussions between device developers and clinicians, academics, and those with lived experience could increase the chances that mental health applications will be considered when manufacturers make changes in the sensors on such devices.

Study design and reporting

This workgroup sought to identify specific elements that are essential for successful digital sensing studies (as measured by significant and reproducible results). There is obviously no “one size fits all” formula; the design and reporting of exploratory studies will differ from hypothesis testing studies (e.g., those focused on identifying how digital sensing features map to specific existing psychological constructs), both of which will differ substantially from the design and reporting of clinical trials (e.g., evaluating whether the incorporation of digital sensing can improve specific clinical outcomes).

As noted in the Introduction, most published digital mental health sensing studies are vastly underpowered for either discovery or replication. There is not yet a sufficient body of literature reporting adequately powered digital sensing investigations that could serve as models in study design for other researchers. In the absence of such evidence, an important step that researchers could take to consider the appropriate scale of digital sensing

studies that they are considering would be to, a priori, model as precisely as possible how digital sensing data types are hypothesized to relate to particular psychological or behavioral constructs, in different types of study samples. Values for these constructs will likely show more variability and change more rapidly in cohorts of individuals with diagnosed mental health disorders compared to population samples. Because of the lower expected effect sizes, researchers designing population studies must plan to investigate large samples for long durations. Additionally, as digital sensing studies generate data for a sizable number of distinct features, the need to correct statistical analyses for multiple testing can substantially affect their power; researchers should consider this factor in deciding on the frequency with which they assess the psychological or behavioral constructs that will be analyzed in relation to digital features.

The problem of statistical power is compounded by heterogeneity in multiple domains that can influence the relationships between sensor-derived features and mental health related targets, such as age, comorbid medical conditions, geographic location (e.g., urban vs. rural, latitude), culture, and socioeconomic status⁸. Additionally, while it is known that sociodemographic factors contribute substantially to the high degree of missing data characteristic of digital sensing studies, there is not yet a principled method to account for such missing data²⁶.

Other medical informatics disciplines have demonstrated the value, for new fields, of establishing repositories for data sharing that can enable a wide community of researchers to apply a multitude of approaches to examine the relationships between measured variables and constructs of interest. An instructive example of such a data repository released for widespread analyses of sensitive medical data is that of the Medical Information Mart for Intensive Care (MIMIC) datasets, which required cross-sector collaboration to develop and have massively enabled innovation in computational medicine^{27,28} including novel imputation²⁹ strategies.

Many research groups have acquired the experience and expertise to identify and adjust for common sources of errors when conducting these studies. Building from a review of digital health studies for the passive monitoring of depression by Angel et al.⁵, the workgroup assembled a list of

Box 3 | Participatory action research playbook overview

Benchmarking and literature review: Conduct a thorough analysis of the target population, including a zip code analysis of needs, historical trauma, and cultural factors that affect the community's trust in research. This information will guide the research team in the development of the research question and hypothesis.

Identifying cultural liaisons: Identify and engage with cultural liaisons who can help the research team understand cultural factors that affect the community's perceptions of research. This will involve identifying individuals who have deep roots in the community and are trusted by community members.

Creating an advisory board: Create an advisory board that includes both internal and external partners. Internal partners should be individuals with authority who are committed to building trust and relevance, such as the Chief of Diversity, Equity, and Inclusion (DEI) within a leading organization. External partners should be community organizers or local NGOs who are already viewed as trusted parties.

Community engagement and outreach strategies: Develop and implement community engagement and outreach strategies that meet the community where they are. This process will involve going to where

people congregate, such as schools or community centers, and using layperson's terms to communicate with clients and stakeholders. The research team should also consider providing culturally adapted translations of all materials.

Prioritizing and aligning objectives: Work with the community to identify common ground and set realistic expectations by articulating the mission statement and objectives of the research. This will help to manage expectations and mitigate unintended consequences and mission creep.

Hypothesis and prototype development: Use feedback from the community to develop a hypothesis and prototype that meet their needs and address the research question. This will involve creating upstream feedback loops and constantly engaging with the community to ensure the research stays on track.

Testing and evaluation: Test the hypothesis and prototype with the community and evaluate the effectiveness of the research. This will involve a constant feedback loop to ensure that the research remains relevant to the community's needs and concerns.

recommendations for variables that researchers should consider when designing studies (Box 2). This list represents a current consensus of the workgroup participants for a *minimum* set of criteria and will evolve as the field matures.

Inconsistent reporting

In considering what steps would enable the field to trust reports of study findings, the group recognized that a lack of standardized formats for reporting results is a source of nonreproducibility of findings beyond those due to differences in platforms or to the issues of power discussed above. It is critical therefore that study reports address issues such as sufficiently characterizing the sensor technology, identifying the reference standard to which the technology is being compared, and documenting nonstandardized measures. As a step toward achieving standardization, the group proposed that reports should include standard checklists of methodology and outcomes and suggested the type of information that these checklists should include. One example of such a checklist is EVIDENCE-MH (EValuatIng connected sENsor teChnologiEs for Mental Health), available online at <https://escholarship.org/uc/item/19c8w68w>. This checklist adapts reporting frameworks developed in other areas of health science^{30,31} to the specific challenges of digital sensing in mental health. The checklist is not intended to duplicate existing guidelines for reporting research, such as CONSORT^{32,33} or STROBE³⁴, but to specify aspects critical to digital sensing studies in mental health. By using EVIDENCE-MH, those preparing, reading, or reviewing personal sensing studies in mental health will be better equipped to determine whether a report has included all of the elements that are crucial for evaluating both the design of a study and its results.

The users' perspective

This workgroup explored ways to involve and engage potential users of digital sensing (people living with mental illness and their families or caregivers, mental healthcare providers, and representatives of diverse communities) in the design, implementation, and reporting of research based on these technologies.

Representation, community trust, and cultural difference in measurement

Realizing the promise of digital sensing technologies in real-world settings requires the integration of different user perspectives from the outset. This point is especially true for research involving consumer

devices, as the potential scale of impact encompasses entire populations. As such, it is imperative that researchers recognize the extent of historical and current inequities that have excluded minoritized communities and other marginalized populations from obtaining benefits from mental health research. These are the groups that are traditionally most impacted by mental health disorders, have had the least access to effective interventions for them, and have been underrepresented at all stages of research and development processes. This lack of representation can make research data biased and incomplete, which in itself reduces the generalizability of study findings. Research based on consumer devices faces a particular equity issue; if mental health digital sensing applications are developed predominantly for the most expensive devices, they may be inaccessible to large segments of the populations that need them. Thus, given the extent of current and historical inequities, ensuring inclusivity in digital health studies is imperative, for example, through the adoption of methodologies that we outline below³⁵.

Digital health studies require distinct recruitment methodologies that reflect their reliance on personal devices with continuous tracking of individual behavior and physiology, often outside a traditional health-care setting. To meaningfully engage diverse users, we propose user centered design and increased use of participatory action research (PAR), where community members are viewed as research partners who collaborate to define the problem, create solutions, test hypotheses, and provide iterative feedback. Toward this goal, we suggest the development of a resource to help mental health researchers using digital technologies to incorporate PAR. An outline of such a "Participatory Action Research Playbook" is shown in Box 3.

Incorporating PAR into digital sensing studies for mental health, for example by engaging Community Advisory Boards (CABs, including individuals with lived experience of mental health conditions) and cultural liaisons to interact with researchers from the initial stages of study design through the final stages of implementation, can greatly enhance the effectiveness of the research and its ultimate impact. For example, during the recruitment process a CAB can participate in community-focused meetings on privacy, which may increase the engagement of participants from historically marginalized communities. This PAR process guides the focus of a study towards designs that are best suited to such communities, for example by adopting protocols aimed at minimizing participant burden (a step that can improve protocol adherence and reduce the need for downstream troubleshooting).

While PAR approaches offer practical ways for acknowledging cultural differences in measurement, intersections of socioeconomic status with other identities, and aligning researcher and community needs, it may not be sufficient to foster trust with participants, especially in the context of events such as data breaches that have occurred at CrisisTextLine³⁶, BetterHelp³⁷, and other digital platforms. Two additional frameworks may be useful for this purpose the Digital Health Equity Framework (DHEF)³⁸ and the Health Social Justice Guide (DHSJG)³⁹. The DHEF provides a key resource to guide exploration of factors influencing health equity considering the multilevel influence of individual, interpersonal, community and societal factors. The DHSJG outlines key areas that should be considered in the development of digital health tools. Additional details on PAR and frameworks useful for enabling such research can be viewed in the working group report online at <https://escholarship.org/uc/item/0p39q718>.

Privacy concerns

A key focus area for future discussion is ethics and methods for privacy and deidentification of data. This is an area in which researchers require input from those in security-related fields, ethics experts, and those with lived experience to better understand what is required, what is possible, and how to build trusting partnerships. While people are willing to share their digital health data for research purposes^{40–42}, constructing precise and protective informed consent is a challenge, given the granularity of data, the extent of data collection, the lack of control researchers have over third party technologies, and the potential for datasets to include people other than consented participants (e.g., in the case of voice or video recordings). The more complex the informed consent, the more complex its implementation, with the need to keep track of what data have been consented for sharing, for how long, with whom, and what effect the consent has on features derived from a given data type. We must accept that there is an inherent tradeoff, in consent forms, between interpretability and accessibility and a level of detail that encompasses all possible risks and ethical challenges. These challenges are especially salient in research conducted with populations that are unable to provide reliable subjective reports of their symptoms, based on their living with certain disorders or disabilities. Because digital sensing may offer unique possibilities for incorporating information about the experiences of such individuals within clinical care, it is imperative that they be included in research studies; the informed consent process must explicitly address privacy issues that are of particular concern to members of these populations.

Conclusion

Research on digital sensing in mental health is a challenging domain, employing user-generated data to improve our understanding of the relationships between mental health disorders and a wide range of behavioral and physiologic variables. To enable reproducible science in the field will require the establishment of standards for obtaining, interpreting, and reporting sensor data, such as the use of the EVIDENCE-MH checklist. Increased support for research infrastructure, e.g., within new centers of excellence, will be essential for investigators to meet such standards. Finally, to develop equitably, the field of digital sensing in mental health must adopt a series of best practices to expand community engagement at all stages of research, for example, adoption of the Participatory Action Research Playbook as an initial baseline for study quality. It is only through concerted efforts in each of these areas that mental health research will achieve the groundbreaking advancements that digital sensing technology has made possible.

Data availability

No datasets were generated or analysed during the current study.

Received: 5 September 2024; Accepted: 14 November 2024;

Published online: 18 December 2024

References

1. Fried, E. Moving forward: how depression heterogeneity hinders progress in treatment and research. *Expert Rev. Neurother.* **17**, 423–425 (2017).
2. Wardenaar, K. J. & de Jonge, P. Diagnostic heterogeneity in psychiatry: towards an empirical solution. *BMC Med.* **11**, 201 (2013).
3. Allsopp, K., Read, J., Corcoran, R. & Kinderman, P. Heterogeneity in psychiatric diagnostic classification. *Psychiatry Res.* **279**, 15–22 (2019).
4. Insel, T. R. Digital phenotyping: a global tool for psychiatry. *World Psychiatry* **17**, 276–277 (2018).
5. Angel, V. D. et al. Digital health tools for the passive monitoring of depression: a systematic review of methods. *Npj Digital Med.* **5**, 3 (2022).
6. Woolley, S. I., Collins, T., Mitchell, J. & Fredericks, D. Investigation of wearable health tracker version updates. *BMJ Heal. Care Inform.* **26**, e100083 (2019).
7. Chikwetu, L. et al. Does deidentification of data from wearable devices give us a false sense of security? A systematic review. *Lancet Digit. Heal.* **5**, e239–e247 (2023).
8. Adler, D. A. et al. Measuring algorithmic bias to analyze the reliability of AI tools that predict depression risk using smartphone sensed-behavioral data. *npj Ment. Heal. Res.* **3**, 17 (2024).
9. NCI-NHGRI Working Group on Replication in Association Studies. et al. Replicating genotype–phenotype associations. *Nature* **447**, 655–660 (2007).
10. Sullivan, P. F. The psychiatric GWAS consortium: big science comes to psychiatry. *Neuron* **68**, 182–186 (2010).
11. Mailman, M. D. et al. The NCBI dbGaP database of genotypes and phenotypes. *Nat. Genet.* **39**, 1181–1186 (2007).
12. Abdellaoui, A., Yengo, L., Verweij, K. J. H. & Visscher, P. M. 15 years of GWAS discovery: realizing the promise. *Am. J. Hum. Genet.* **110**, 179–194 (2023).
13. Kainec, K. A. et al. Evaluating accuracy in five commercial sleep-tracking devices compared to research-grade actigraphy and polysomnography. *Sensors* **24**, 635 (2024).
14. Shcherbina, A. et al. Accuracy in wrist-worn, sensor-based measurements of heart rate and energy expenditure in a diverse cohort. *J. Personalized Med.* **7**, 3 (2017).
15. IEEE. IEEE Standard for open mobile health data—representation of metadata, sleep, and physical activity measures. *IEEE 1752*, 1–24 (2021).
16. Sim, I., Carini, S. & Zaman, M. *IEEE P1752 Open Mobile Health*. <https://sagroups.ieee.org/1752/> (2024).
17. Society, D. M. *DATAcc | Homepage - DATAcc by DiMe*. <https://datacc.dimesociety.org/> (2024).
18. Estrin, D. & Sim, I. Open mHealth architecture: an engine for health care innovation. *Science* **330**, 759–760 (2010).
19. mHealth Data Interoperability. *Open mHealth*. <https://www.openmhealth.org/> (2024).
20. Parnell, L. D. et al. BioStar: An online question & answer resource for the bioinformatics community. *PLoS Comput. Biol.* **7**, e1002216 (2011).
21. NSF. *Smart and Connected Communities (S&CC) | NSF—National Science Foundation*. <https://new.nsf.gov/funding/opportunities/smart-connected-communities-scc> (2024).
22. Bent, B. et al. The digital biomarker discovery pipeline: an open-source software platform for the development of digital biomarkers using mHealth and wearables data. *J. Clin. Transl. Sci.* **5**, e19 (2020).
23. Wilkinson, M. D. et al. The FAIR guiding principles for scientific data management and stewardship. *Sci. Data* **3**, 160018 (2016).
24. Lin, D. et al. The TRUST principles for digital repositories. *Sci. Data* **7**, 144 (2020).
25. Goldsack, J. C. et al. Verification, analytical validation, and clinical validation (V3): the foundation of determining fit-for-purpose for

biometric monitoring technologies (BioMeTs). *npj Digit. Med.* **3**, 55 (2020).

26. Kiang, M. V. et al. Sociodemographic characteristics of missing data in digital phenotyping. *Sci. Rep.* **11**, 15408 (2021).
27. Mark, R. *Secondary Analysis of Electronic Health Records* 1st edn, Vol. 427 (2016).
28. Johnson, A. E. W. et al. MIMIC-IV, a freely accessible electronic health record dataset. *Sci. Data* **10**, 1 (2023).
29. Shadbahr, T. et al. The impact of imputation quality on machine learning classifiers for datasets with missing values. *Commun. Med.* **3**, 139 (2023).
30. Manta, C. et al. EVIDENCE publication checklist for studies evaluating connected sensor technologies: explanation and elaboration. *Digit. Biomark.* **5**, 127–147 (2021).
31. Coravos, A. et al. Modernizing and designing evaluation frameworks for connected sensor technologies in medicine. *NPJ Digit. Med.* **3**, 37 (2020).
32. Schulz, K. F., Altman, D. G., Moher, D. & Group, C. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMJ* **340**, c332 (2010).
33. Moher, D. et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ* **340**, c869 (2010).
34. Elm et al. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet* **370**, 1453–1457 (2007).
35. Pratap, A. et al. Indicators of retention in remote digital health studies: a cross-study evaluation of 100,000 participants. *npj Digit. Med.* **3**, 21 (2020).
36. Levine, A. S. *Suicide Hotline Shares Data With for-Profit Spinoff, Raising Ethical Questions*. <https://www.politico.com/news/2022/01/28/suicide-hotline-silicon-valley-privacy-d> (2022).
37. Bajak, F. *BetterHelp Shared Users' Sensitive Health Data, FTC Says*. <https://apnews.com/article/betterhelp-ftc-health-data-privacy-befca40bb873661d1f8986bb75d8df07> (2023).
38. Richardson, S., Lawrence, K., Schoenthaler, A. M. & Mann, D. A framework for digital health equity. *NPJ Digit. Med.* **5**, 119 (2022).
39. Figueroa, C. A. et al. Applying the digital health social justice guide. *Front. Digit. Heal.* **4**, 807886 (2022).
40. Gupta, R. et al. Consumer views on privacy protections and sharing of personal digital health information. *JAMA Netw. Open* **6**, e231305 (2023).
41. Seltzer, E. et al. Patients' willingness to share digital health and non-health data for research: a cross-sectional study. *BMC Méd. Inform. Decis. Mak.* **19**, 157 (2019).
42. Shandhi, M. M. H. et al. Assessment of ownership of smart devices and the acceptability of digital health data sharing. *npj Digit. Med.* **7**, 44 (2024).
43. Mohr, D. C., Zhang, M. & Schueller, S. M. Personal sensing: understanding mental health using ubiquitous sensors and machine learning. *Annu Rev. Clin. Psycho.* **13**, 23–47 (2017).

Acknowledgements

The 2023 Workshop on Advancing the Utility of Digital Sensing Tools for Mental Health Research was sponsored by Wellcome, NIMH and the University of California Los Angeles (UCLA) Depression Grand Challenge. It was organized by Nelson Freimer (UCLA), Simona Carini (University of California San Francisco, UCSF), Tanzeem Choudhury (Cornell Tech), Matthew Hotopf (King's College London, KCL), and David Mohr (Northwestern University). We gratefully acknowledge the individuals who attended and contributed to the workshop in the various working groups. The Digital Infrastructure workgroup included the following individuals: Mami Bartlett (Apple), Matt Bianchi (Apple), Richard Dobson (KCL), Omer Inan (Georgia Institute of Technology), Rony Krell (Optum Labs at UnitedHealthCare), Santosh Kumar (University of Memphis), Ian Marcus (Food and Drug Administration), Michelle Popowitz (UCLA), and Gwydion

Williams (Wellcome). The Data Flow workgroup included: Tim Althoff (University of Washington), Justin Baker (McLean Hospital), Jakob Bardram (Technical University of Denmark), Matthew Bietz (University of California Irvine), Jeff Cochran (Otsuka Pharmaceutical), Greg Farber (NIMH), Vaibhav Narayan (Davos Alzheimer's Collaborative), Loes Olde Loohuis (UCLA), and Elizabeth Shriberg (Ellipsis Health). The Study Design & Reporting workgroup included: Brunilda Balliu (UCLA), Raphe Bernier (Apple), Sandra Bucci (The University of Manchester), Anind Dey (University of Washington), Jonathan Flint (UCLA), Suzanne Gage (Wellcome), Adam Haim (NIMH), Daniel McDuff (Google/University of Washington), Margaretta Nyilas, Mashfiqui Rabbi (Optum Labs at UnitedHealthCare), Alastair van Heerden (Human Science Research Council, South Africa), and Til Wykes (KCL). The Users' Perspective workgroup included: Ricardo Araya (KCL), Jessie Bakker (Digital Medicine Society), Jae Canas (UCLA), Michelle Craske (UCLA), Anne Dahlem (UCLA), Carlos Lopez-Jaramillo (Universidad de Antioquia), Inacio Silva (Federal University of Pelotas, Brazil), Kunmi Sobowale (UCLA), Brandon Staglin (One Mind), Kelechi Ubozoh (Kelechi Ubozoh Consulting), and Yuki Zhang (Otsuka Pharmaceuticals). We would like to extend gratitude to those who provided feedback and insights during the public comment period on the initial workshop working group reports. Special thanks to John Torous (Beth Israel Deaconess Medical Center), Vincent Leonard Holstein (McLean Hospital), Imogen Bell (Orygen), and Eve Rosenfeld (Stanford University).

Author contributions

SA: Writing—Original Draft, Writing—Review and Editing, Visualization. DS: Writing—Review and Editing. CD: Writing—Review and Editing. AA: Writing—Review and Editing. SC: Writing—Review and Editing. JD: Writing—Review and Editing. MH: Writing—Review and Editing. DM: Writing—Review and Editing. AB: Writing—Review and Editing, Supervision, Project administration, Funding acquisition. NF: Conceptualization, Writing—Review and Editing, Supervision, Project administration, Funding acquisition.

Competing interests

AA is a clinical advisor and has stock options with BigHealth and BeMe. The remaining co-authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Nelson B. Freimer.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024