Background paper

Digital Footprints and Real-Time Data in the Context of Health Research: Ethical Issues

Author: Anita Ho, PhD, MPH[1]

Outline of Paper

I. Introduction

II. Purpose

III. Overview

IV. Sources, Applications, and Use of Digital Footprint and Real-Time Data
- Wearables and sensor data
- Consumer data
- Scraping internet searches and AI prompts
- Social media listening

V. Health Research Using Digital Footprints and Real-Time Data: Key Ethical issues
- Data quality, bias, and equity
- Data sovereignty and commodification
- Consent, privacy, and datafication
- Ethical governance

VI. Conclusions

The background paper key themes and questions are summarised in this video:

[1] This background paper is the result of a collaborative effort. I thank members of the GFBR Steering Committee and Planning Committee for their constructive and valuable feedback on earlier drafts.

I. Introduction

Digital technologies are transforming the landscape of data gathering and health research involving human participants globally. Traditionally, health data refer to information collected from patients, including diagnoses, medications, treatment plans, test results, clinical trial data, or encounters in clinical research or the formal health care setting. Sources of health data may include medical and electronic health records, genomic data, research data, and administrative data (e.g., insurance claims), which are identifiable and generally protected by various regulatory regimes. De-identified population data on vital statistics, health surveys, and disease registries on clinical and outcomes data for a predefined patient population complement individual health data to help researchers and health systems understand health trends in various communities.

As digital technologies become ubiquitous, more data than ever are now being collected about humans across multiple directly and indirectly health-related indicators and determinants, blurring the boundaries of health data. In addition to the digitalisation of clinical and laboratory data collected in the formal health care setting, individuals also generate digital footprints,[2] which are the unique trail of data individuals create, either actively or passively, when using the internet or connected devices. These digital footprints are often real-time data,[3] which is information available for processing and analysis immediately after it is generated or collected. For example, wearable devices, smartphones, social media platforms, consumer technologies, internet searches, and artificial intelligence (AI) prompts are generating vast quantities of multi-modal data about individuals' behaviours, movements, physiological states, and social interactions.

Digital footprints and streams of real-time data are generally collected by non-health sectors, especially commercial entities, for other primary purposes. Nonetheless, these data may become health-relevant when linked with other variables[4] or when hypothesised to be proxy indicators of health outcomes. They are increasingly leveraged in health research, such as biomedical research that aims at generating new knowledge around biological mechanisms, treatment modalities, and healthcare system matters to prevent, diagnose, and treat diseases. They are also utilised in public health research that addresses population health and health system concerns, as well as in social research that explores the social, cultural, and behavioural factors shaping health outcomes. The COVID-19 pandemic, rapid advances in AI to manage data, and the expansion of learning health systems have further accelerated interest in using digital footprints at scale for health-related research. While there is high hope in research communities that digital footprints can enhance our ability to understand disease patterns, predict risk, monitor interventions, formulate nudge policies, and inform public

[2] IBM. What is a digital footprint? https://www.ibm.com/think/topics/digital-footprint (accessed 3 March, 2026).
[3] IBM. What is real-time data? https://www.ibm.com/think/topics/real-time-data (accessed 3 March, 2026).
[4] Stuart R, Sergio L, Effy V, George C, Tiwonge M, Nezerith C, Jerome S, Walter J, Keymanthri M. Private commercial companies sharing health-relevant consumer data with health researchers in sub-Saharan Africa: an ethical exploration. Policy Stud. 2024 Sep 19:10.1080/01442872.2024.2403506. doi: 10.1080/01442872.2024.2403506. Epub ahead of print. PMID: 40857480; PMCID: PMC12356098.

health responses,[5] these novel data sources may lead to new research approaches that raise distinct ethical issues.

Focusing on the context of low- and middle-income countries (LMICs), this year's Global Forum on Bioethics in Research (GFBR) will explore the ethical collection and use of real-time data and digital footprints in health-related research, particularly around how this research should be designed, conducted, and governed. Proponents see the potential for digital footprints and real-time data to transform health-related research in LMICs, whether by offering novel and more accessible approaches to data collection or by enabling data generation where other means of data collection (e.g., electronic health records) are unavailable or inoperable. Critics worry about representativeness and accuracy of these data, and question whether and how we might decide to promote real-time data gathering techniques over traditional approaches, and under what circumstances. These tensions highlight the importance of having a good understanding of the potential ethical challenges that may arise in the use of digital footprints and real-time data for health-related research, and that we have LMIC-centered approaches and frameworks for working through these challenges. Promoting scientifically sound and ethically robust research through sharing and using high-quality data is critical to maximising the potential benefits of health-related research and respecting relevant communities, citizens, and research participants.

Broad ethical principles of respecting the rights and interests of individual participants, minimising harm to research participants while promoting participants' and communities' health and well-being, and promoting distributive justice regarding benefits and burdens have guided many biomedical research ethics frameworks.[6] For research conducted in low-resource settings, the Council for International Organisations of Medical Sciences (CIOMS) has highlighted the importance of fair collaboration, community engagement, and protections for populations that may experience various forms of vulnerability.[7] Yet, as an increasing amount of health-related research is data-related research, which may

---

[5] Burgess R, Dolan E, Poon N, Jenneson V, Pontin F, Sivill T, Morris M, Skatova A. Harnessing digital footprint data for population health: a discussion on collaboration, challenges and opportunities in the UK. BMJ Health Care Inform. 2024 Sep 28;31(1):e101119. doi: 10.1136/bmjhci-2024-101119. PMID: 39343444; PMCID: PMC11448216.

[6] International Ethical Guidelines for Health-related Research Involving Humans, Fourth Edition. Geneva. Council for International Organizations of Medical Sciences (CIOMS); 2016. Available at: https://cioms.ch/publications/product/international-ethical-guidelines-for-health-related-research-involving-humans/; World Medical Association. WMA Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects. First adopted in June 1964, and last amended in October 2024. Available at: https://www.wma.net/policies-post/wma-declaration-of-helsinki/; National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research [Internet]. Washington: U.S. Government Printing Office; 1978. Available at: https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/read-the-belmont-report/index.html; South Africa. Regulations relating to research with human participants. Pretoria: Government Printer; 2014 Sep 19. (National Health Act, 2003 (Act No. 61 of 2003)). Available at: https://www.gov.za/sites/default/files/gcis_document/201409/38000rg10268gon719.pdf

[7] International Ethical Guidelines for Health-related Research Involving Humans, Fourth Edition. Geneva. Council for International Organizations of Medical Sciences (CIOMS); 2016. https://cioms.ch/publications/product/international-ethical-guidelines-for-health-related-research-involving-humans/.

not fit the narrow traditional biomedical model of research involving humans, this background paper asks: **what are various ethical complexities that arise from using digital footprints and real-time data in health research?** And as more sectors and actors with divergent priorities are involved in the expanding digital footprints ecosystems, **how well do existing ethical frameworks address these issues, and what other ethical oversight mechanisms may be needed to tackle new concerns when data are collected and used across diverse and non-homogeneous ecosystems?** For example, what uses of personal and/or health data warrant research ethics review when digital footprints from a range of health **and** non-health sources are increasingly used for health research? As we seek appropriate ethical governance of digital footprints in research that is proportional to the possible threats to autonomy, individual welfare, public good, or other social values, there are also prior questions of how we can ascertain the quality and utility of such data. And how may we best address ethical issues around inclusive research and reciprocity of benefits in the context of ongoing power asymmetry in the global order, especially given the increasing commodification of data? And what are the key ethical concerns around individuals and communities often being unaware of how their digital footprints are being repurposed for health research? While people's right to not participate in health research may need to be balanced with their positive ethical duties to contribute to public health advancement via potentially useful data research, in the context of health research using digital footprints, questions remain how the evolving multi-sectoral research enterprise may meet societal expectations around not only voluntary participation, but also values of reciprocity and non-exploitation.[8]

To address these rapidly changing data practices and associated ethical concerns, **a central goal of this GFBR meeting is to map the current ethical debates and tensions around the collection and use of digital footprints and real-time data in health research in LMICs**. As evolving data practices increasingly involve multi-sectoral stakeholders who have not been traditionally part of health research ecosystems (e.g., software developers, corporations), but now play a role in contributing to or eroding trustworthiness of data-sharing practices, **an overarching theme of this meeting is the ethical governance for using digital footprints and real-time data in health research in LMICs**. This meeting encourages multidisciplinary discussions on how we can share and use data *well* in a complex, cross-sectoral setting. Recognising ongoing resource disparity and digital divides within and across regions that may have different health priorities, the GFBR meeting will identify existing use cases of digital footprints for health research, explore current ethical review models, identify unresolved ethical questions or stubborn problems, and propose governance solutions that are ethically justified, pragmatically operationalisable, and which may allow us to radically reimagine how we define and ethically approach the use of digital footprints and real-time data in health research.

[8] Carter P, Laurie GT, Dixon-Woods M. The social licence for research: why *care.data* ran into trouble. Journal of Medical Ethics 2015;41:404-409.

II. Purpose

This paper is being published with the 2026 GFBR call for participants and presenters. It provides background information and further details on the forum topic and scope. Proposed presentations may relate to the themes in this paper or other issues that present **ethical challenges related to the use of digital footprints and real-time data for human health research**. While there may be ambiguity regarding whether real-time data are collected and used in research versus a public health context, we welcome discussions around various ethical issues and oversight implications of these grey areas. The forum's focus is on health research rather than public health *per se*. However, we welcome presentations that explore whether or how the rapidly evolving data landscape may warrant new definitions of health research.

Submissions should focus on health research in the context of LMICs. Examples from high-income countries (HICs) are also considered if they show relevance to LMIC settings. Proposals could address (but are not limited to) one or more of the questions in Section V, 'Health Research Using Digital Footprints and Real-Time Data: Key Ethical Issues.'

GFBR is seeking four types of presentation proposals:

1. **Real-life case studies** that explore ethical dimensions of health research using human digital footprints, including how the following dimensions interact with core research ethics principles and shape our understanding of ethical research practice:
    - **Different types of data sources** (e.g., internet search history, social media posts, fitness trackers, industry data of purchasing patterns, geolocation data)
    - **Different types of health applications** (e.g., for individual health management, health behaviour monitoring, clinical applications, public health surveillance, environmental risks monitoring, or remote monitoring in clinical trials)
    - **Different (presumed) data ownerships** (e.g., private, public, commercial)
    - **Different types of health conditions or phenomena** (e.g., infectious diseases, sexual and reproductive health, mental health, non-communicable and chronic illnesses, etc.)
    - **Different types of users generating or using the data for health research** (e.g., clinicians, patients, healthy research participants, technology/AI developers, health researchers)
    - **Different geographical areas where digital footprints are collected, processed, and interpreted** (e.g., LMICs vs HICs)
    - **Different methods of data collection, types of permissions, and consent by different sectors** (e.g., public vs private vs civil society/third sector, user agreements, terms and conditions)
    - **Different granularities of the data and associated quality issues** (aggregate data and metadata vs identifiable or personal data)
    - **Different ethical governance models across sectors involving parties with variable training and understanding of ethics** (e.g., health

researchers, health care professionals, AI developers, commercial software engineers, data scientists)

2. **Conceptual papers** that define, clarify, and analyse key moral terms to address ethical dimensions of using digital footprints and real-time data for human health research. Examples might include the meaning and moral work done in the current context by concepts such as autonomy, social value, public benefit, consent, privacy, public interest, social license, trust and trustworthiness.

3. **Normative papers** that provide in-depth ethical analysis or explore unresolved ethical questions related to the topic. Examples may include, among others:
    - Debates around public/private information
    - Data ownership
    - Duties to contribute to public health advancements
    - Respect for participants through consent vs other governance models
    - Right to benefit from outcomes of sharing data
    - Considerations around equity – e.g., concerns around perpetuating dominance of some groups' representation and further marginalisation of others.

4. **Policy-related papers** that address ethical issues that arise from governance or policy matters related to the topic or explore the translation of ethics and value-based decision-making into policy. Examples:
    - Ethical frameworks for institutional, national, regional or international regulation, guidelines, or policies that address the collection and use of digital footprints and real-time data in health research
    - Research ethics review or governance of evolving real-time data practices
    - Data regimes and their effectiveness in promoting ethical health research using digital footprints
    - Challenges of delivering good governance in a cross-sectoral health research environment
    - Papers that examine and reflect upon the role of ethics in policymaking for health research that uses digital footprints and real-time data
    - Papers that discuss models and examples of successes and failures of ethical translation into policymaking
    - Papers that identify the necessary steps to improve the uptake of ethics into policy in this field.

III. Overview

Data collection across interconnected personal, clinical, and public health domains is foundational to modern health care delivery and health research worldwide. Reliable and high-quality health-related data can enable informed clinical and policy decision-making, and support accountability and equity in health systems. These data are essential for understanding disease burdens and their distribution, developing treatment plans, allocating scarce resources, and designing equitable domestic and international

public health interventions. As we have learned in epidemic outbreaks, including the COVID-19 pandemic, early detection of communicable diseases is crucial to organise and allocate the required or scarce health resources, to control the spread of the disease, and to avoid or mitigate further contamination.[9]

Traditional sources for health-related data such as clinical and laboratory data, research data, self-reported surveys, health registries, pharmacovigilance databases, and administrative records are crucial for population health monitoring and health research. Nonetheless, they are limited in their ability to swiftly respond to emerging health trends and emergencies. Public health monitoring data for health research have practical limitations due to resource constraints, slow or infrequent survey cycles, incomplete participation, recall and reporting biases, and time lag between data collection and reporting, although actual experiences and capacities vary among different regions and countries. Participation in cohort studies and epidemiological surveys across the globe is declining, especially in disadvantaged and hard-to-reach populations, who bear a disproportionate burden of communicable and non-communicable diseases (NCDs). Patient symptom reports and clinical data often only represent a specific time frame and may miss gradual changes that require longitudinal monitoring. In addition, there are technical limitations such as the lack of electronic medical records in lower-resource settings and the absence of interoperability across different platforms that render data-sharing for health research challenging.

With rapid advances in digital technologies, researchers and health decision-makers are pursuing other types of data and data sources to promote real-time monitoring and efficient data-sharing. The expanding availability of ‘smart’ and social data, the widespread penetration of mobile phones and internet usage, and the increasing power of machine learning to analyse large datasets have fueled the growth of digital epidemiology. Epidemiology, very broadly speaking, is the understanding of disease patterns and their causes, and to use this understanding to mitigate and prevent diseases as well as to promote health. Digital epidemiology is the use of digital data, which may or may not have been originally generated for research or epidemiological purposes, and may or may not be directly related to health in its most familiar sense, to supplement traditional data to better understand and improve public health.[10]

There is high hope in health research communities that these new methods and expanded data sources will have less practical and technical challenges and can provide granular and large-scale insights to enhance health research,[11] although some of these challenges remain across digital data. These evolving digital data practices involve diverse approaches and multiple stakeholders across very different data regimes, with varying ethical standards, awareness and levels of training (if any). As a

[9] Pilipiec P, Samsten I, Bota A. Surveillance of communicable diseases using social media: A systematic review. PLoS One. 2023 Feb 24;18(2):e0282101. doi: 10.1371/journal.pone.0282101. PMID: 36827297; PMCID: PMC9956027.
[10] Salathé M. Digital epidemiology: what is it, and where is it going? Life Sci Soc Policy. 2018 Jan 4;14(1):1. doi: 10.1186/s40504-017-0065-7. PMID: 29302758; PMCID: PMC5754279.
[11] Licinio J, Wong ML. Digital footprints as a new translational approach for mental health care: a commentary. Discov Ment Health. 2023 Jan 30;3(1):5. doi: 10.1007/s44192-023-00032-7. PMID: 37861744; PMCID: PMC10501006.

result, questions abound whether existing ethical approaches and frameworks are adequate in governing these new research methods and data practices in LMICs, particularly when examined across global contexts marked by incommensurate systems of governance and data infrastructure, and inequalities of access and control.

This background paper provides an overview of the multi-dimensional ethical challenges and questions related to the use of digital footprints and real-time data in health research involving humans in LMICs. It begins with a brief survey of common types of sources, applications, and use of real-time data in health research, highlighting the expanding variety of actors in this evolving space and salient ethical considerations for these respective data sources and associated research methods. It is then followed by a discussion of four ethical themes of particular interest to this year's GFBR:

1) data quality, bias, and equity
2) data sovereignty and commodification
3) consent, privacy, and datafication
4) ethical governance.

Within each theme, we raise broad and specific ethical questions around health research using digital footprints and real-time data. By also highlighting the interrelationships between these themes and questions, this background paper situates the rapidly expanding real-time data practices to broader global historical and structural practices. It serves as a vehicle to help GFBR participants:

1) explore the novel ethical challenges that arise in the use of this data for health-related research
2) consider whether, in light of these novel challenges, new ethical guidelines and governance approaches or frameworks are necessary, and, if so,
3) deliberate on what form they might take.

IV. Sources, Applications, and Use of Digital Footprint and Real-Time Data

Four key 'smart' data sources and their applications in health research are of particular interest to this forum because of their expanding use in health research and associated ethical concerns:

1) wearables and biosensors
2) consumer data
3) internet searches and AI prompt scraping
4) social media listening.

As we will see, while currently most of the data generated by these sources begin with non-health related activities repurposed for health research (e.g., geolocation data repurposed for infectious disease research), some applications may have been designed or deployed for the specific goal of informing research (e.g., digital biomarkers research studies, physical activity tracking studies). In the age of big data, it is also notable that these different data sources may be merged with other forms of data for

more granular analysis or as part of a larger health research process. This section will provide a brief overview of each data source being discussed, particularly around how it is used for health-related purposes, how it may be used in health research, and pertinent ethical questions that may arise out of using such real-time data for health research.

1) Wearables and sensor data

The global wearable technology market size is projected to grow from USD 92.90 billion in 2025 to USD 229.97 billion by 2033.[12] Wearables are body-worn, internet-enabled electronic devices that monitor activity and physiological signals. They can generate a broad spectrum of health, wellness, and physical activity data outside of the clinic setting to provide information about personal wellbeing to the consumer.[13] Biomedical sensors are increasingly integrated into clothing and accessories or directly applied to the skin. Other sensors such as implantable devices have also been introduced. For example, implantable glucose monitoring has been commercially available in South Africa since 2016 for adults with diabetes.[14] Depending on the specific functionality, sensors can provide continuous, real-time measurements of physiological and biochemical parameters such as heart rate, body temperature, sleep patterns, stress and glucose levels, and hydration status.[15] Wearable sensors that can provide longitudinal remote monitoring and/or real-time predictive analysis are frequently proposed as non-invasive and user-friendly tools to facilitate personal health management and data collection tools that can enhance users' health outcomes and health research.[16] Newer Internet of Things (IoT) devices, which are the integrated network of smart and connected devices that can communicate and share real-time

[12] Grand View Research. Wearable Technology Market (2026 - 2033). https://www.grandviewresearch.com/industry-analysis/wearable-technology-market (accessed 3 March 2026).

[13] Devine JK, Schwartz LP, Hursh SR. Technical, Regulatory, Economic, and Trust Issues Preventing Successful Integration of Sensors into the Mainstream Consumer Wearables Market. Sensors (Basel). 2022 Apr 2;22(7):2731. doi: 10.3390/s22072731. PMID: 35408345; PMCID: PMC9002880.

[14] Tweden KS, Deiss D, Rastogi R, Addaguduru S, Kaufman FR. Longitudinal Analysis of Real-World Performance of an Implantable Continuous Glucose Sensor over Multiple Sensor Insertion and Removal Cycles. Diabetes Technol Ther. 2020 May;22(5):422-427. doi: 10.1089/dia.2019.0342. PMID: 31697182; PMCID: PMC7196365.

[15] Vo DK, Trinh KTL. Advances in Wearable Biosensors for Healthcare: Current Trends, Applications, and Future Perspectives. Biosensors (Basel). 2024 Nov 18;14(11):560. doi: 10.3390/bios14110560. PMID: 39590019; PMCID: PMC11592256.

[16] Mukherjee MD, Gupta P, Kumari V, Rana I, Jindal D, Sagar N, et al. Wearable biosensors in modern healthcare: Emerging trends and practical applications. Talanta Open. 2025 Dec. https://doi.org/10.1016/j.talo.2025.100486; Smith AA, Li R, Tse ZTH. Reshaping healthcare with wearable biosensors. Sci Rep. 2023 Mar 27;13(1):4998. doi: 10.1038/s41598-022-26951-z. PMID: 36973262; PMCID: PMC10043012.

data with minimum human intervention,[17] may provide efficient ways for clinicians and health researchers to receive, measure, and analyse health data.[18]

Wearable sensor devices, especially when AI-powered, go beyond simply recording discrete events at various times. Many of these applications are passive technologies. Once activated, they continuously collect, store, and analyse different forms of data simultaneously, without the user's active prompt or involvement. Fitness and wellness devices that track users' steps, heart rate, and duration of activities are some examples. Other clinical applications of AI wearables and biosensors include devices that can provide real-time diagnostic information or predictive forecasting of a future condition. For example, wearable sensors in smart watches can now analyse ECG (Electrocardiogram) signals to detect atrial fibrillation and other arrhythmias suggesting heart failure. Research on wearable biosensor technologies in India, United States, and China also shows promise in AI-assisted epilepsy management, offering real-time, continuous monitoring and early seizure prediction/detection.[19]

In LMICs, wearable and biosensor use is less prevalent than in HICs due to a range of factors, including high costs of devices and mobile data access, network coverage, limited system integration of wearable devices into existing infrastructures, data privacy concerns,[20] and other cultural considerations.[21] Nonetheless, pilot studies or donor-funded projects targeting specific NCDs are accelerating their utilisation. As 85% of premature deaths from NCDs such as cardiovascular diseases, cancers, respiratory diseases, and diabetes occur in LMICs,[22] health systems with limited resources may particularly benefit from wearables and biosensors to enhance screening, prevention, and monitoring of NCDs.[23] Wearables may also supplement limited institutional records and continuity between clinic visits by giving individuals more information for their own health management. In Ghana, continuous glucose monitoring devices have demonstrated potential in improving diabetes care among young people with Type 1

[17] Mandari HE. Acceptance of Internet of Things in Developing Countries: An Empirical Study Using Value-Based Adoption Model. Int J Technol Hum Interact [Internet]. 2022 Jan 1;18(1):1-19. https://doi.org/10.4018/IJTHI.300288.

[18] Herman Eliewaha Mandari, Acceptance of Internet of Things in Developing Countries:, International Journal of Technology and Human Interaction, Volume 18, Issue 1, 2022, ISSN 1548-3908, https://doi.org/10.4018/IJTHI.300288.

[19] Aziz S, A M Ali A, Aslam H, Ul Ain N, Tariq A, Sohail Z, Murtaza S, Mahmood HI, Wazeer MI, Murtaza F, Abd-Alrazaq A, Alsahli M, Damseh R, AlSaad R, Shahzad T, Ahmed A, Sheikh J. Wearable Artificial Intelligence for Epilepsy: Scoping Review. J Med Internet Res. 2025 Oct 31;27:e73593. doi: 10.2196/73593. PMID: 41172347; PMCID: PMC12578435.

[20] Jamil S, Mohammadnezhad M, Abdulrahim A, Muhammad F, Khan HTA. Managing Diabetes One Step at a Time in Low- and Middle-Income Countries: The Promise of Wearable Devices. Chronic Dis Transl Med. 2025 Aug 12;11(4):279-283. doi: 10.1002/cdt3.70018. PMID: 41341735; PMCID: PMC12670968.

[21] Sachdeva M, Dugerdil A, Flahault A, Carrara V. Are Wearables Effective in LMICs? Public Health Rev. 2025 May 9;46:1607940. doi: 10.3389/phrs.2025.1607940. PMID: 40416068; PMCID: PMC12098041.

[22] Pan American Health Organization. Non-Communicable Diseases. https://www.paho.org/en/topics/noncommunicable-diseases (accessed 5 February 2026).

[23] Sachdeva M, Dugerdil A, Flahault A, Carrara V. Are Wearables Effective in LMICs? Public Health Rev. 2025 May 9;46:1607940. doi: 10.3389/phrs.2025.1607940. PMID: 40416068; PMCID: PMC12098041.

diabetes, with participants reporting improved blood sugar control and increased awareness of lifestyle impacts.[24]

Wearable and sensor devices may also help to expand monitoring or research capacity within a 'lean' implementation environment (e.g., the use of pulse oximeters to monitor hospitalised patients with COVID-19 in Vietnam).[25] They are increasingly employed in research for disease monitoring, drug discovery, and pollutant detection.[26] These technologies may enhance health research in LMICs that requires ongoing monitoring of health-related behaviours and physiological information. They may reduce participant burden by minimising in-person research sessions for physiological data recording. Health researchers studying individuals managing chronic or degenerative conditions can use wearable sensors to track symptoms, health behaviours, treatment adherence, and functional status continuously in a real-world setting, enabling more informed longitudinal studies of population-level health.[27] For environmental health research, citizen scientists in African cities such as Cape Town, Lagos, and Accra have also used running and low-cost wearable air quality sensors to generate evidence for decision makers.[28]

Early data suggest that residents in LMICs are interested in using wearable devices for health management and health research. Studies in Uganda,[29] Kenya,[30] and Burkina Faso[31] have all shown high acceptance of using these devices for research purposes.

---

[24] Jamil S, Mohammadnezhad M, Abdulrahim A, Muhammad F, Khan HTA. Managing Diabetes One Step at a Time in Low- and Middle-Income Countries: The Promise of Wearable Devices. Chronic Dis Transl Med. 2025 Aug 12;11(4):279-283. doi: 10.1002/cdt3.70018. PMID: 41341735; PMCID: PMC12670968.
[25] Chau NVV, Trung TN, Khanh PNQK et al. Wearable devices for remote monitoring of hospitalized patients with COVID-19 in Vietnam [version 1; peer review: 1 approved]. Wellcome Open Res 2022, 7:257 (https://doi.org/10.12688/wellcomeopenres.18026.1).
[26] Bhalla N, Jolly P, Formisano N, Estrela P. Introduction to biosensors. Essays Biochem. 2016 Jun 30;60(1):1-8. doi: 10.1042/EBC20150001. PMID: 27365030; PMCID: PMC4986445.
[27] Muurling M, de Boer C, Kozak R, Religa D, Koychev I, Verheij H, Nies VJM, Duyndam A, Sood M, Fröhlich H, Hannesdottir K, Erdemli G, Lucivero F, Lancaster C, Hinds C, Stravopoulos TG, Nikolopoulos S, Kompatsiaris I, Manyakov NV, Owens AP, Narayan VA, Aarsland D, Visser PJ; RADAR-AD Consortium. Remote monitoring technologies in Alzheimer's disease: design of the RADAR-AD study. Alzheimers Res Ther. 2021 Apr 23;13(1):89. doi: 10.1186/s13195-021-00825-4. PMID: 33892789; PMCID: PMC8063580.
[28] Clean Air Fund. 2023. Young runners use wearable tech to capture air quality data in African cities. https://www.cleanairfund.org/case-study/young-runners-use-wearable-tech-to-capture-air-quality-data-in-african-cities/ (accessed 3 March 2026).
[29] Swahn MH, Gittner KB, Lyons MJ, Nielsen K, Mobley K, Culbreth R, Palmier J, Johnson NE, Matte M, Nabulya A. Advancing mHealth Research in Low-Resource Settings: Young Women's Insights and Implementation Challenges with Wearable Smartwatch Devices in Uganda. Sensors. 2024; 24(17):5591. https://doi.org/10.3390/s24175591.
[30] Kwaro D, Mendt S, Okoth J, Munga S, Gunga HC, Heim ZH, Matzke I, Bunker A, Barteit S, Maggioni MA. Acceptability and feasibility of research grade wearables for monitoring heat stress in Kenyan farmers. NPJ Digit Med. 2025 May 7;8(1):255. doi: 10.1038/s41746-025-01601-6. PMID: 40335647; PMCID: PMC12059195.
[31] Huhn S, Matzke I, Koch M, Gunga HC, Maggioni MA, Sié A, Boudo V, Ouedraogo WA, Compaoré G, Bunker A, Sauerborn R, Bärnighausen T, Barteit S. Using wearable devices to generate real-world, individual-level data in rural, low-resource contexts in Burkina Faso, Africa: A case study. Front Public Health. 2022 Sep 30;10:972177. doi: 10.3389/fpubh.2022.972177. PMID: 36249225; PMCID: PMC9561896.

Nonetheless, technical issues remain, and these studies mostly focused on participants' experience with the device itself rather than their perspectives on data-sharing. The gaps in these studies raise questions of what ethical obligations cross-sectoral stakeholders have to promote user understanding of how wearable sensor data may be used for health research, as well as the benefits and risks of such research.

It is also worth noting that, in most LMICs, wearable devices that track users' mental and/or physical well-being but are not marketed as medical devices are generally not subjected to regulatory scrutiny or approval, such that the quality and safety of these devices and their data practices are difficult to ascertain.[32] This raises questions of what ethical responsibilities researchers may have to validate wearable data before using them to draw conclusions that could influence health policies or medical practice, and what implications this may have on device companies that control the data. As companies may hold divergent interests from research enterprises, there are also questions of what cross-sectoral ethical review or governance approaches are appropriate for using wearable sensor data for health research.

2) Consumer data

Digital consumer data, for our purpose, include data collected about consumers as well as data from mobile devices and consumer applications. Examples include mobile app engagement data, shopping and dining data, and geolocation data. In LMICs, with the ubiquity of mobile phones and increasing connectivity, expanding sources of consumer data have been repurposed to complement traditional epidemiological data sources to model population movement, disaster response, disease transmission, and environmental events to help design and evaluate various health interventions.[33]

Some consumer devices and applications target specific health or wellness issues. Pregnancy applications, which dominate the worldwide health-related direct-to-consumer (DTC) market, combine physiological data, self-reported symptoms, and other behavioural information (e.g., diet, exercise) to predict and inform expectant users of progress and fetal growth.[34] In LMICs, there is increasing hope that pregnancy apps that require minimal cost and equipment can eventually be tailored for early detection and management of high-risk conditions such as preeclampsia,[35] which is a leading cause of global maternal morbidity and mortality. Researchers can explore whether

[32] Ho A. *Live Like Nobody is Watching: Relational Autonomy in the Age of Artificial Intelligence Health Monitoring*. New York: Oxford University Press. 2023.
[33] Suhag A, Burgess R, Skatova A. Shopping Data for Population Health Surveillance: Opportunities, Challenges, and Future Directions. J Med Internet Res. 2025 Aug 6;27:e75720. doi: 10.2196/75720. PMID: 40769214; PMCID: PMC12327962.
[34] Ho A. Artificial Intelligence and Emerging Technologies. In: Wendy Rogers, Stacy Carter, Catherine Mills, Jackie Leach Scully and Vikki Entwistle. *Routledge Handbook of Feminist Bioethics*. Routledge. 2022. p. 291-307.
[35] Ibrahim AM, Jahanfar S. Effectiveness and equity of mHealth apps for preeclampsia management in LMICs: A rapid review protocol. PLoS One. 2024 Nov 13;19(11):e0313655. doi: 10.1371/journal.pone.0313655. PMID: 39536068; PMCID: PMC11559984.

information from these apps can help to fill gaps in access to health care solutions and reduce adverse maternal-fetal events in low-resource settings.[36]

Other consumer data serve as digital footprints for health research. For example, cellphone geolocation data is increasingly utilised for real-time and longitudinal tracking of travel or commuting patterns, outdoor activity habits, visits to different shopping and food outlets, and exposure to environmental pollutants. These secondary data may support health research by shedding light on how behaviour and environment interact in affecting health outcomes.[37] The geographic distributions of location data may also enhance research to design and evaluate health interventions for specific communities, or allocate community health workers and other resources to impoverished areas. During infectious disease outbreaks and climate events (e.g., drought, flooding, extreme heat waves), cellphone mobility data have been used to model impact areas and assess effectiveness of different interventions[38] and strategies[39] (e.g., self-quarantine, get tested). For example, during the 2014-2016 Ebola epidemic in West Africa, researchers designed and evaluated a smart phone application for contact tracing. While the study team reported substantial challenges of using an app in the epidemic context, the study demonstrated that it was possible to implement mobile health in this emergency setting.[40] A few years later, during the COVID-19 pandemic, phone apps with geo-sensing and proximity tracking technologies were developed in Kenya to track patients' locations and alert people with SMS instructions when they might have come near an infected individual.[41] In Malawi, researchers used mobile phone data to identify the occurrence and size of potential mass gatherings to understand COVID-19 transmission spread to previously unaffected areas.[42]

---

[36] Jonas SM, Deserno TM, Buhimschi CS, Makin J, Choma MA, Buhimschi IA. Smartphone-based diagnostic for preeclampsia: an mHealth solution for administering the Congo Red Dot (CRD) test in settings with limited resources. J Am Med Inform Assoc. 2016 Jan;23(1):166-73. doi: 10.1093/jamia/ocv015. Epub 2015 May 29. PMID: 26026158; PMCID: PMC7814923.

[37] Chambers E. 2025. Smartphone Location Data Shows Promise for Public Health Research. https://medicine.wsu.edu/news/2025/04/21/smartphone-location-data-research/ (accessed 3 March 2026).

[38] Emish M, Kelani Z, Hassani M, Young SD. A Mobile Health Application Using Geolocation for Behavioral Activity Tracking. Sensors (Basel). 2023 Sep 15;23(18):7917. doi: 10.3390/s23187917. PMID: 37765972; PMCID: PMC10537358.

[39] Buckee CO, Balsari S, Chan J, Crosas M, Dominici F, Gasser U, Grad YH, Grenfell B, Halloran ME, Kraemer MUG, Lipsitch M, Metcalf CJE, Meyers LA, Perkins TA, Santillana M, Scarpino SV, Viboud C, Wesolowski A, Schroeder A. Aggregated mobility data could help fight COVID-19. Science. 2020 Apr 10;368(6487):145-146. doi: 10.1126/science.abb8021. Epub 2020 Mar 23. PMID: 32205458.

[40] Danquah LO, Hasham N, MacFarlane M, Conteh FE, Momoh F, Tedesco AA, Jambai A, Ross DA, Weiss HA. Use of a mobile application for Ebola contact tracing and monitoring in northern Sierra Leone: a proof-of-concept study. BMC Infect Dis. 2019 Sep 18;19(1):810. doi: 10.1186/s12879-019-4354-z. PMID: 31533659; PMCID: PMC6749711.

[41] Müller A, Cau A, Muhammed S, Abdullahi O, Hayward A, Nsanzimana S, Lester R. Digital mHealth and Virtual Care Use During COVID-19 in 4 Countries: Rapid Landscape Review. JMIR Form Res. 2022 Nov 30;6(11):e26041. doi: 10.2196/26041. PMID: 34932498; PMCID: PMC9714961.

[42] Green D, Moszczynski M, Asbah S, et al. Using mobile phone data for epidemic response in low resource settings—A case study of COVID-19 in Malawi. Data & Policy. 2021;3:e19. doi:10.1017/dap.2021.14.

Expanding digital footprints, including purchasing, dining, loyalty card, and mobile metadata, also provide secondary data that may enhance health research. Retail data on groceries, alcohol, and tobacco can bypass individuals' reports and provide objective,[43] real-time measures of consumption for health research and epidemiologic understanding of various lifestyle-related diseases.[44] During the COVID-19 pandemic, these consumption data also facilitated research on whether the pandemic had other indirect health behavioural consequences.[45] Consumer data from over-the-counter (OTC) medication (e.g., cough medicine, allergy medications) and smart thermometer readings may also help researchers track patterns of illness across the year and around various regions faster than traditional clinical reporting. Online entertainment and video consumption may support research around the impact of these services on various communities' physical and mental health.[46] Furthermore, smartcard transit data in LMICs may inform research on the equity of health-seeking behaviours across different transport users.[47]

Nonetheless, as consumer digital footprints that are collected for heterogenous purposes are increasingly being leveraged to make inferences about people's health, there are questions of whether existing research ethics frameworks are adequate to evaluate the use of these data for health research purposes, including potential biases, methodological limits, and pitfalls of this type of research.[48] New uses of consumer data push us to reflect on the meanings of "health" and "health research". These data could be used to research people's general wellbeing habits, which might in turn be utilised to inform a nudge agenda in ways that are not necessarily in the health interests of citizens. Depending on whether this is categorised as "health research," it could fall outside of ethical governance regimes. As consumer data may reveal detailed information or make speculations about individuals, there are also questions of how researchers can address the inherent privacy tension both between the individual and corporations as well as between individuals and government. For example, pregnancy app developers and researchers may originally aim to use these technologies and associated data to promote and study reproductive health and prenatal care needs.

---

[43] Skatova A. Overcoming biases of individual level shopping history data in health research. NPJ Digit Med. 2024 Sep 30;7(1):264. doi: 10.1038/s41746-024-01231-4. PMID: 39349949; PMCID: PMC11442457.

[44] Jenneson VL, Pontin F, Greenwood DC, Clarke GP, Morris MA. A systematic review of supermarket automated electronic sales data for population dietary surveillance. Nutr Rev. 2022 May 9;80(6):1711-1722. doi: 10.1093/nutrit/nuab089. PMID: 34757399; PMCID: PMC9086796.

[45] Lee BP, Dodge JL, Leventhal A, Terrault NA. Retail Alcohol and Tobacco Sales During COVID-19. Ann Intern Med. 2021 Jul;174(7):1027-1029. doi: 10.7326/M20-7271. Epub 2021 Mar 2. PMID: 33646843; PMCID: PMC7983313.

[46] Downs S. 2025. Measuring the Health Impact of Consumer Products and Services. https://www.milbank.org/2025/01/consumer-products-and-services-are-a-key-driver-of-health/ (accessed 28 March 2026).

[47] Du F, Wang J, Liu Y, Zhou Z, Jin H. Equity in Health-Seeking Behavior of Groups Using Different Transportations. Int J Environ Res Public Health. 2022 Feb 27;19(5):2765. Doi: 10.3390/ijerph19052765. PMID: 35270458; PMCID: PMC8910309.

[48] Olteanu A, Castillo C, Diaz F, Kıcıman E. Social Data: Biases, Methodological Pitfalls, and Ethical Boundaries. Front Big Data. 2019 Jul 11;2:13. Doi: 10.3389/fdata.2019.00013. PMID: 33693336; PMCID: PMC7931947.

However, data that suggest a person's pregnancy has terminated may subject the monitored user to legal risk in jurisdictions that have restrictions on reproductive health interventions. To promote responsible use of consumer digital footprints in health research, we welcome presentations that explore meanings of "health" and "health research" in the expanding use of consumer footprints to help inform appropriate ethical and regulatory responses. We also encourage submissions that explore various ways that public engagement may help to align public expectations and values with these data research practices.

3) Scraping internet searches and AI prompts

As digital and internet infrastructure continues to improve in various parts of the world, the public is increasingly engaging in online and AI queries for information, guidance, and entertainment, including for health-related information. Globally, Google Trends can show spikes in search terms (e.g., flu remedies), filtered by date or location, that may reveal clues to new outbreaks. During the COVID-19 pandemic, when the novel, sudden, and highly contagious global health event led to home isolation and limited clinic visits for non-urgent conditions across the world, online health information seeking (OHIS) became the main channel for many people seeking health information. Google's online hub for search trends related to COVID-19 was used to help researchers and health systems explore underreported symptoms and predict regional spikes in diagnoses.[49]

Nonetheless, there is a disparity of OHIS within and between countries,[50] raising ethical questions of whether uncritical reliance on OHIS for health research may lead to biased results. In China, one study revealed that middle-aged and older adults were less likely to engage in OHIS than younger adults, and individuals with high school and college attainments were more inclined to engage in OHIS than those with a middle school education or below.[51] A 2021 cross-sectional, nationally representative study in Turkey revealed that 48.6% of participants used the internet as a source of health-related information.[52] In comparison, in Ghana, a 2019 cross-sectional study showed that while 85.8% of the population reported using the internet, only 31.4% engaged in OHIS, with

[49] Stephens-Davidowitz S. 2025. Google Searches Can Help Us Find Emerging Covid-19 Outbreaks. The New York Times. 2020 Apr 5. https://www.nytimes.com/2020/04/05/opinion/coronavirus-google-searches.html.

[50] Jia X, Pang Y, Liu LS. Online Health Information Seeking Behavior: A Systematic Review. Healthcare (Basel). 2021 Dec 16;9(12):1740. doi: 10.3390/healthcare9121740. PMID: 34946466; PMCID: PMC8701665.

[51] Zhao L, Liu T. Exploring generational and educational disparities in online health information seeking in China: The moderating role of internet use. Digit Health. 2025 Nov 4;11:20552076251393378. doi: 10.1177/20552076251393378. PMID: 41200540; PMCID: PMC12586879.

[52] Özkan S, Tüzün H, Dikmen AU, Aksakal NB, Çalışkan D, Taşçı Ö, Güneş SC. The Relationship Between Health Literacy Level and Media Used as a Source of Health-Related Information. Health Lit Res Pract. 2021 Apr;5(2):e109-e117. doi: 10.3928/24748307-20210330-01. Epub 2021 May 10. PMID: 34251938; PMCID: PMC8241229.

computer and internet experience as well as higher economic status being positive predictors of OHIS.[53]

The rapid development of large language models (LLMs) in recent years has further expanded online health inquiries. LLMs are deep learning AI systems trained on a vast amount of data. These models use unsupervised learning to understand and generate human-like text, code, and other content. They are the foundation models for AI chatbots, some of which have multilingual capabilities that enhance global use. As online searches and LLMs allow anonymous queries, and AI chatbots are interactive, people who do not have adequate access to or do not feel comfortable asking clinicians certain questions may turn to these platforms for assessment and guidance,[54] including mental health support.[55] Some chatbots (e.g., ChatGPT) are publicly available services that require low bandwidth, and are gradually reaching an expanding proportion of the global population.[56] Analysing internet searches and AI prompts in LMICs may allow research into unmet health needs, diagnostic uncertainty, and health anxieties. Research on what search and query prompts are most used, what online sources are frequently accessed by various populations and regions, how users interact with these sources, and the comparative accuracy among information sources may shed light on how internet and AI queries may affect users' knowledge, competence, involvement in health decision-making, and utilisation of health services.[57] Such research may also help health systems to design strategies or programs to identify and combat infodemics,[58] when too much information, including false or misleading information during a disease outbreak, may cause confusion and risk-taking behaviours that can harm health.

However, it is noteworthy that users of these platforms may not be informed that their queries regarding personal health matters can be repurposed for health research, or what will happen to their data when services or company operations change. Corporations that own these online and AI platforms have tremendous control over the continuity and functionality of these platforms and the data they generate. While

---

[53] Nangsangna RD, da-Costa Vroom F. Factors influencing online health information seeking behaviour among patients in Kwahu West Municipal, Nkawkaw, Ghana. Online J Public Health Inform. 2019 Sep 19;11(2):e13. doi: 10.5210/ojphi.v11i2.10141. PMID: 31632607; PMCID: PMC6788904.

[54] Bean AM, Payne RE, Parsons G, Kirk HR, Ciro J, Mosquera-Gómez R, Hincapié M S, Ekanayaka AS, Tarassenko L, Rocher L, Mahdi A. Reliability of LLMs as medical assistants for the general public: a randomized preregistered study. Nat Med. 2026 Feb;32(2):609-615. doi: 10.1038/s41591-025-04074-y. Epub 2026 Feb 9. PMID: 41663592; PMCID: PMC12920132.

[55] Ho A, Perry J. What We Owe Those Who Chat Woe: A Relational Lens for Mental Health Apps. Am J Bioeth. 2023 Oct;23(10):77-80. doi: 10.1080/15265161.2023.2250306. Epub 2023 Oct 9. PMID: 37812122.

[56] Wang X, Sanders HM, Liu Y, Seang K, Tran BX, Atanasov AG, Qiu Y, Tang S, Car J, Wang YX, Wong TY, Tham YC, Chung KC. ChatGPT: promise and challenges for deployment in low- and middle-income countries. Lancet Reg Health West Pac. 2023 Sep 15;41:100905. doi: 10.1016/j.lanwpc.2023.100905. PMID: 37731897; PMCID: PMC10507635.

[57] Alma Taya D, Chuang YC. Internet use for health information, health service utilization, and quality of care in the U.S. BMC Health Serv Res. 2025 May 8;25(1):659. doi: 10.1186/s12913-025-12807-5. PMID: 40340831; PMCID: PMC12060370.

[58] World Health Organization. Infodemic. https://www.who.int/health-topics/infodemic#tab=tab_1 (accessed 5 February 2026).

companies may provide standard and often verbose “Terms and Conditions” agreements for consumers using these services, they are generally pre-determined by the companies, and individual users may have little control over their data except not to engage in any of the service. Changes in ownership model and corporate decisions around their search engines and AI products may also affect how users interact with these technologies, and what may happen to their data. Salient examples include Open AI’s move from being non-profit to for-profit and its decision to shut down their AI video generator, Google’s decision to prioritise AI overview for internet searches, and chatbot developers’ decisions to tweak their model functionality. For researchers, there is an inherent vulnerability in depending on digital footprints for health research, as corporate decisions may unilaterally affect data availability and reliability, since user behaviors may alter due to these changes. These intersecting issues highlight cross-sectoral challenges in determining appropriate ethical governance and gaining public trust for using internet searches and AI queries for health research.[59] These examples also raise questions about the ethical responsibilities of diverse actors in this domain to demonstrate their trustworthiness in how they generate, process, share and use data from such sources.

4) Social media listening

As the global population gains increasing access to the internet and mobile phone connections, especially in South Asia, Africa, and Latin America,[60] social media platforms have experienced rapid growth. 92.6% of all adults aged 18 and above around the world use some form of social media, with Facebook, YouTube, and Instagram being the most widely used platforms globally.[61] These open and social online channels allow users to interact in real-time or asynchronously.[62] Users can share personal and health-related information and beliefs in general-purpose social media (e.g., Facebook) or in health-related social networks (e.g., DailyStrength or MedHelp).[63] Social media platforms can connect with large audiences and user involvement is typically higher than conventional web-based health interventions.[64]

---

[59] Carter P, Laurie GT, Dixon-Woods M. The social licence for research: why care.data ran into trouble. J Med Ethics. 2015 May;41(5):404-9. doi: 10.1136/medethics-2014-102374. Epub 2015 Jan 23. PMID: 25617016; PMCID: PMC4431337.

[60] World Population Review. Social Media Users by Country 2026. https://worldpopulationreview.com/country-rankings/social-media-users-by-country (accessed 5 February 2026).

[61] Datareportal. Global Social Media Statistics. https://datareportal.com/social-media-users#:~:text=And%20for%20additional%20context%2C%20the,View%20fullsize (accessed 5 February 2026).

[62] Carr CT, Hayes RA. Social media: defining, developing, and divining. Atlantic Journal of Communication 2015; 23:46–65.

[63] Paul, M. J., Sarker, A., Brownstein, J. S., Nikfarjam, A., Scotch, M., Smith, K. L., & Gonzalez, G. 2016. Social media mining for public health monitoring and surveillance. In Pacific Symposium on Biocomputing 2016, PSB 2016 (pp. 468-479). World Scientific Publishing Co. Pte Ltd.

[64] Kanchan S, Gaidhane A. Social Media Role and Its Impact on Public Health: A Narrative Review. Cureus. 2023 Jan 13;15(1):e33737. doi: 10.7759/cureus.33737. PMID: 36793805; PMCID: PMC9925030.

Social media listening involves tracking and analysing publicly available posts, comments, or interactions from diverse social media sources. X (formerly known as Twitter) is used by 29.2% of internet users aged 16 and above, and is the most cited source for studies exploring social media textual mining for the purpose of the surveillance and prediction of communicable diseases.[65] Social media listening can be useful for research on public sentiment, misinformation, and emerging health issues. It may promote understanding of how communities perceive risks and interventions, and has the potential to improve the public health surveillance system.[66] During the COVID-19 pandemic, researchers used social media listening to analyse emotions and opinions in various regions surrounding vaccine safety[67] and various forms of public health restrictions.[68] Social media listening also allows social network analysis by mapping connections and influencers within these platforms that may include large populations.

In LMICs, social media listening may enhance research into marginalised or hard-to-reach populations that have phone and internet connection but have less access to clinical services and research opportunities. As publicly available social media data can bypass many logistical hurdles associated with traditional approaches of data collection,[69] it may provide health researchers a more rapidly accessible source of real-world data on population and patient experiences.[70] When integrated with other forms of data, social media listening may help researchers and health systems identify, categorise, and characterise social narratives to inform evidence-based interventions.[71]

Nonetheless, questions abound what privacy expectations identifiable users have regarding their social media posts that may consist of personal and sensitive information. Moreover, available data on social media demographics show that active users tend to be younger, women, more highly educated and less acutely ill or

---

[65] Pilipiec P, Samsten I, Bota A. Surveillance of communicable diseases using social media: A systematic review. PLoS One. 2023 Feb 24;18(2):e0282101. doi: 10.1371/journal.pone.0282101. PMID: 36827297; PMCID: PMC9956027.

[66] Pilipiec P, Samsten I, Bota A. Surveillance of communicable diseases using social media: A systematic review. PLoS One. 2023 Feb 24;18(2):e0282101. doi: 10.1371/journal.pone.0282101. PMID: 36827297; PMCID: PMC9956027.

[67] Bucci LM, Lamprianou S, Gesualdo F, Tozzi AE, Ghalayini T, Sahinovic I, Pal S. A social media intervention for communicating vaccine safety in low- and middle-income countries: protocol for a pilot study. Front Public Health. 2023 Dec 7;11:1248949. doi: 10.3389/fpubh.2023.1248949. PMID: 38145079; PMCID: PMC10748494.

[68] Russell SN, Rao-Graham L, McNaughton M. Mining social media data to inform public health policies: a sentiment analysis case study. Rev Panam Salud Publica. 2024 Dec 16;48:e79. doi: 10.26633/RPSP.2024.79. PMID: 39687240; PMCID: PMC11648203.

[69] McDonald L, Malcolm B, Ramagopalan S, Syrad H. Real-world data and the patient perspective: the PROmise of social media? BMC Med. 2019 Jan 16;17(1):11. doi: 10.1186/s12916-018-1247-8. PMID: 30646913; PMCID: PMC6334434.

[70] Street J, Farrell L. Analysis of social media. In: Facey KM, Ploug Hansen H, Single ANV, editors. Patient involvement in health technology assessment [Internet]. Singapore: Springer; 2017. 10.1007/978-981-10-4068-9.

[71] World Health Organization. 2025. Social listening in infodemic management for public health emergencies: Guidance on ethical considerations. https://iris.who.int/server/api/core/bitstreams/e16284bc-abfe-42e5-9444-aeab1a174b0d/content (accessed 5 February 2026).

functionally impaired, raising issues of external validity or representativeness of using social media listening for inclusive health research.[72] Since users of social media platforms may not be representative of the broader population (e.g., Reddit users are mostly young men), and patterns of online expression vary across cultures, age groups, and political environments, overreliance on social media listening in health research may risk overrepresenting vocal minorities or misrepresenting silent or offline communities. Moreover, changes in corporate structure or decisions (e.g., the sale of Twitter to Elon Musk) may affect what content or users may be amplified or excluded, which may in turn impact which and how users engage on social media platforms. It is also worth noting that a recent study shows that chatter on social media about global events comes from 20% bots and 80% humans,[73] raising additional validity questions of using social media posts for health research. In addition, as we will see in the next section, there are methodological and quality concerns around using proxy measures (e.g., social media posts) that may have inherent biases and have not been validated in health research to infer health status (e.g., psychosis risk).[74]

V. Health Research Using Digital Footprints and Real-Time Data: Key Ethical issues

The rapidly expanding sources of digital footprints and real-time data have the potential to enhance researchers' ability to explore health-related behaviours, develop new clinical applications, and strengthen health system surveillance and planning efforts in LMICs. However, as we saw in various data source examples, digital footprints often originate and circulate across diverse domains, with a widening range of stakeholders and actors being engaged in the evolving ecosystem. In addition to traditional clinician scientists, health care practitioners and academic researchers, corporations, software engineers, and data scientists with divergent interests and variable training in research ethics are now involved in the digital footprints lifecycle.

As new stakeholders and actors occupy the diverse digital data environment, interconnected ethical questions abound around the quality of digital footprints and the knowledge generated through such research in LMICs. How *useful* may data-intensive health research that uses digital footprints be in promoting wellbeing, fairness, and justice, particularly in the context of data commodification and ongoing power and resource asymmetry in global health? In the increasingly complex landscape in which data are gathered and used, how may we determine the appropriate levels of transparency, privacy protection, accessibility, inclusiveness, and benefit-sharing in

---

[72] McDonald L, Malcolm B, Ramagopalan S, Syrad H. Real-world data and the patient perspective: the PROmise of social media? BMC Med. 2019 Jan 16;17(1):11. doi: 10.1186/s12916-018-1247-8. PMID: 30646913; PMCID: PMC6334434.

[73] Ng LHX, Carley KM. A global comparison of social media bot and human characteristics. Sci Rep. 2025 Mar 31;15(1):10973. doi: 10.1038/s41598-025-96372-1. PMID: 40164745; PMCID: PMC11958817.

[74] Olteanu A, Castillo C, Diaz F, Kıcıman E. Social Data: Biases, Methodological Pitfalls, and Ethical Boundaries. Front Big Data. 2019 Jul 11;2:13. doi: 10.3389/fdata.2019.00013. PMID: 33693336; PMCID: PMC7931947. Readers may also be interested in the July 2025 Special Issue of Asian Bioethics Review, which included articles on a range of ethical issues on social listening. Available at https://link.springer.com/journal/41649/volumes-and-issues/17-3.

cross-sectoral data research to help reduce vulnerability and promote fairness? In navigating the involvement of multiple sectors, how do we prevent multiple systems of review and approvals that may result in unnecessary duplication of ethical effort, a lack of proportionality in governance, and ultimately a risk that scientifically sound health research might be thwarted?

In this changing data research landscape, the overarching question for this year's GFBR is whether existing ethical frameworks and guidelines are adequate or appropriate for these evolving approaches involving data collection, utilisation, and sharing, and whether additional or new governance structures are necessary for these data-focused research approaches. In this section, we introduce key ethical issues and questions of particular interest to this year's forum. We invite exploration of what ethical models and governance mechanisms may best promote ethical health research using digital footprints and real-time data. We also welcome submissions that raise other questions related to these themes:

1) Data quality, bias, and equity
2) Data sovereignty and commodification
3) Consent, privacy, and datafication
4) Ethical governance.

1) Data quality, bias, and equity

The expanding utilisation of digital footprints for health research in LMICs raises important ethical questions partly because significant uncertainties remain about the quality and reliability of heterogeneous real-time data. Digital footprints are generated across sectors for different primary purposes and operating under different technical standards. While many countries maintain regulatory frameworks governing medical devices, these regulations typically apply only to software or AI tools that function explicitly as medical devices, such as diagnostic algorithms. In contrast, consumer technologies that collect physiological signals or behavioural data often fall outside formal regulatory oversight.[75] As a result, datasets derived from these technologies may be used or incorporated into health research without clear evidence that they meet established standards for measurement of validity, reliability, or verifiability, the last of which is a particular concern for black box AI models.

CIOMS has established the conditions under which biological samples and health-related data may be used in human research.[76] These guidelines emphasise that studies involving human participants or their data must demonstrate scientific validity and social value. Fulfilling these guidelines in the context of digital footprints for health

[75] Devine JK, Schwartz LP, Hursh SR. Technical, Regulatory, Economic, and Trust Issues Preventing Successful Integration of Sensors into the Mainstream Consumer Wearables Market. Sensors. 2022; 22(7):2731. https://doi.org/10.3390/s22072731.

[76] International Ethical Guidelines for Health-related Research Involving Humans, Fourth Edition. Geneva. Council for International Organizations of Medical Sciences (CIOMS); 2016. https://cioms.ch/publications/product/international-ethical-guidelines-for-health-related-research-involving-humans/.

research is challenging because there is currently limited guidance on how to evaluate the validity of consumer products and consumer-generated data outside of the controlled laboratory or clinical environments. Digital data from consumer products are increasingly used for research, even though their software or models are not designed by the research team.[77] This includes research that prospectively uses consumer devices (e.g., wearables) as tools to collect data for a specific study. Commercial digital data sources frequently contain methodological limitations such as sampling bias, measurement error, and uncertain validity, which are often non-transparent to researchers or health systems. When internal testing is conducted by technology companies, details regarding sampling rates, device sensitivity, algorithmic design, computational methods, or error rates across different populations are rarely disclosed, as companies often treat their training data sources and other methodological information as proprietary.

This lack of transparency contrasts with expectations in conventional biomedical research, where investigators are typically required to report recruitment and data collection strategies, demographic characteristics, and sampling procedures to demonstrate that data are representative and appropriate for analysis. Health researchers relying on commercial products or datasets may be unable to verify whether data were collected or processed consistently or whether error rates differ across demographic groups. These uncertainties raise concerns regarding the scientific validity, accountability, and potential clinical value of digital data,[78] particularly when such data are used to guide health interventions or policy decisions affecting populations in LMICs. There are questions of which ethical obligations research communities must demonstrate to uphold the trustworthiness and integrity of digital data used in health research, and what level of access to raw or underlying data controlled by private companies is necessary to enable transparent and accountable research practices.[79] Quality data-driven research relies in turn on quality metadata. "Metadata" is literally data about data, and a robust metadata system is essential to the quality and functioning of any dataset to be used for health research. Concerns about low-level data quality can have profound implications for the quality of metadata, raising additional worries about data provenance, renewed privacy risks, and problems with attribution of credit for new research findings.

Concerns about data quality intersect closely with ethical issues of bias and health equity, since measurement errors or unrepresentative datasets can disproportionately

---

[77] Lucivero F, Prainsack B. The lifestylisation of healthcare? 'Consumer genomics' and mobile health as technologies for healthy lifestyle. Appl Transl Genom. 2015 Feb 7;4:44-9. doi: 10.1016/j.atg.2015.02.001. PMID: 26937349; PMCID: PMC4745358.

[78] Devine JK, Schwartz LP, Hursh SR. Technical, Regulatory, Economic, and Trust Issues Preventing Successful Integration of Sensors into the Mainstream Consumer Wearables Market. Sensors. 2022; 22(7):2731. https://doi.org/10.3390/s22072731.

[79] Salauddin N. Evaluating the role of metadata standards in enhancing data discoverability and interoperability in academic digital repositories. *Digital Library Perspectives.* 2026 Vol. ahead-of-print No. ahead-of-print. https://doi.org/10.1108/DLP-07-2025-0111; Bernardi FA, Alves D, Crepaldi N, Yamada DB, Lima VC, Rijo R. Data Quality in Health Research: Integrative Literature Review. J Med Internet Res. 2023 Oct 31;25:e41446. doi: 10.2196/41446. PMID: 37906223; PMCID: PMC10646672.

affect already marginalised populations. Individuals who engage with digital platforms may differ systematically from those who do not.[80] For example, as we saw earlier, there is disparity with OHIS and AI usage. Moreover, demographic and socio-economic biases may exist in consumer or shopping data, limiting the representativeness of these datasets.[81] When digital footprints collected under such conditions are used in health research, they may produce or reinforce inaccuracies that disproportionately affect underserved populations. Such data biases may be even more consequential in LMICs, if imported technologies are adopted in research studies without extensive local validation due to limited resources or presumption of model generalisability. Subjecting people in LMICs to devices and interventions that have questionable data quality or safety, especially in the service of commercial interests, instrumentalises humans and violates their dignity. If low-quality data research is used to inform clinical care or health policies, it may also cause additional harm to those who are already socially vulnerable.

Global digital divides and broader disparities further highlight these ethical concerns around access, benefit-sharing, and justice, as not all individuals or communities have equal ability to generate or benefit from digital data. Across the world, there are inequitable digital and research infrastructures that see resources concentrated in HICs and not equitably shared with LMICs that are contributing vital data to health research and scientific knowledge. Consequently, the patterns embedded in these datasets may reflect the epidemiological, social, and healthcare conditions of HICs rather than those of LMICs. At the same time, attempts to supplement datasets with digital footprints from LMICs (e.g., internet searches, social media activity, or mobile phone usage) are shaped by unequal access to connectivity and digital devices. For example, qualitative research conducted in rural India suggests that women's access to mobile phones may be limited by cost constraints, limited digital skills, and patriarchal gender norms.[82] Health research using these data may therefore disproportionately reflect the experiences of wealthier, urban, younger, and often male users, while marginalising the perspectives of rural, older, or socioeconomically disadvantaged groups and thus their equal access to research benefits. Worse yet, health research based on biased digital datasets may misrepresent and produce inaccurate results for underrepresented populations that, when applied, may harm these groups. These issues raise questions of what ethically appropriate strategies research communities can adopt to prevent reinforcing existing digital divides and global inequalities and to promote inclusive and equitable data collection and benefit-sharing practices.

---

[80] Burgess R, Boyd A, Davis OS, Millard LA, Mumme M, Robertson S, Skinner A, Xiao Z, Skatova A. Linking digital footprint data into longitudinal population studies. Int J Popul Data Sci. 2025 Jun 3;10(1):2946. doi: 10.23889/ijpds.v10i1.2946. PMID: 40463363; PMCID: PMC12132027.

[81] Skatova A. Overcoming biases of individual level shopping history data in health research. NPJ Digit Med. 2024 Sep 30;7(1):264. doi: 10.1038/s41746-024-01231-4. PMID: 39349949; PMCID: PMC11442457.

[82] Scott K, Shinde A, Ummer O, Yadav S, Sharma M, Purty N, Jairath A, Chamberlain S, LeFevre AE. Freedom within a cage: how patriarchal gender norms limit women's use of mobile phones in rural central India. BMJ Glob Health. 2021 Sep;6(Suppl 5):e005596. doi: 10.1136/bmjgh-2021-005596. PMID: 34551901; PMCID: PMC8461288.

- What are the ethically salient harms associated with conducting health research using biased, incomplete, or unrepresentative digital footprint datasets?
- What strategies or mechanisms can be used to mitigate the risk of perpetuating or exacerbating existing health inequities, particularly when underlying datasets reflect structural disparities in access to technology or healthcare in LMICs? How can inclusion be promoted?
- What methods, processes, or tools can help researchers assess data quality across diverse digital ecosystems, particularly when health research relies on data collected through heterogeneous methods, purposes, and institutional standards? Who else, besides researchers, should be involved in assessing bias and equity issues (e.g., research ethics committees, interested communities, scientific committees, etc.)?
- What are the ethically justifiable limits of proprietary control over digital footprints when such data are used for health research, especially when commercial interests may conflict with values such as transparency, scientific validity, public accountability, and the promotion of population health?

2) Data sovereignty and commodification

Issues around inequity in the previous section highlights how global health research and data infrastructure continue to be shaped by profound power asymmetry,[83] exacerbated by increasing commercialisation of personal and health-related data. With that backdrop, data sovereignty is increasingly recognised as a key ethical issue in health research in LMICs.[84]

Data sovereignty is the principle that individuals, communities, and nations have the right to govern the collection and use of their data as part of their "national asset." It is partly about respect, collective self-determination, accountability, and justice. While international data-sharing for reciprocal benefits may be the ethical ideal, data sovereignty is particularly important where data systems and the analytical process are beset by historical legacies of colonialism and present-day injustices.[85] For example, while Africa bears the largest burden of communicable and non-communicable diseases globally, it contributes only about 2% of global health research output.[86] This is partly

[83] Sekalala S, Andanda P, Chatikobo T. Regulating digital health in the Global South: critical and decolonial approaches. International Journal of Law in Context. 2026;22(1):1-10. doi:10.1017/S1744552325100335.
[84] Sassmannshausen F. 2026. As WHO Debates Global AI Regulation, States Clash Over 'Data Sovereignty'. https://healthpolicy-watch.news/who-debates-global-ai-rules/ (accessed 5 March 2026).
[85] Sekalala S, Andanda P, Chatikobo T. Regulating digital health in the Global South: critical and decolonial approaches. International Journal of Law in Context. 2026;22(1):1-10. doi:10.1017/S1744552325100335.
[86] Adebamowo CA, Callier S, Akintola S, Maduka O, Jegede A, Arima C, Ogundiran T, Adebamowo SN; BridgELSI Project as part of the DS-I Africa Consortium. The promise of data science for health research in Africa. Nat Commun. 2023 Sep 29;14(1):6084. doi: 10.1038/s41467-023-41809-2. PMID: 37770478; PMCID: PMC10539491.

because of inaccessibility and low systematic maintenance of medical data in LMICs,[87] but also partly because of a lack of investment by major funders or benefit-sharing by HICs. In 2023, LMICs received less than 0.4% of all direct grants for biomedical research by major funders.[88] And even though data from LMICs are often made readily available to trainees and researchers in HICs and are used to generate global evidence, local research teams and health systems may lack access to analytic tools or decision-making power over how data are interpreted, applied, and shared, resulting in a relative dearth of publications of treatment outcomes research and health data registries in LMICs.[89]

Similar issues around data colonialism exist with digital footprints and real-time data for health research. Many LMICs, particularly in urban areas, have experienced high mobile phone penetration. This has facilitated "leapfrogging" in generating novel forms of real-time data at unprecedented volumes. Digital wearable devices are also increasingly available in LMICs, often through partnerships with international institutions or donor programs in HICs. Nonetheless, questions remain how much control LMICs have regarding local data in international health research and whether LMICs will have affordable access to the research knowledge created from their real-time data. Donor programs from HICs may set agendas with minimal local involvement based on donor priorities rather than local needs,[90] and participants may have minimal understanding or control over how their data collected by novel technologies developed in HICs may be stored and used. Data extracted from users in LMICs are often transferred, analysed, and processed in donor countries and by researchers in those regions, further advancing their knowledge and careers while limiting local control over data governance around secondary uses, potentially reinforcing epistemic injustice and exploitation at scale.

Data commodification, whereby multinational corporations are turning human digital footprints into a new monetisable asset, further exacerbates ethical concerns of direct exploitation. Companies may collect and store data in their platforms and assert proprietary claims over the data. These issues are particularly salient in the context of cross-border data extraction. Multinational corporations may harvest data from LMICs to train AI models or develop digital products in HICs that are then sold back to LMICs for

---

[87] Abdul-Rahman T, Ghosh S, Lukman L, Bamigbade GB, Oladipo OV, Amarachi OR, Olanrewaju OF, Toluwalashe S, Awuah WA, Aborode AT, Lizano-Jubert I, Audah KA, Teslyk TP. Inaccessibility and low maintenance of medical data archive in low-middle income countries: Mystery behind public health statistics and measures. J Infect Public Health. 2023 Oct;16(10):1556-1561. doi: 10.1016/j.jiph.2023.07.001. Epub 2023 Jul 7. PMID: 37566992.

[88] World Health Organization. Investments on grants for biomedical research by funder, type of grant, health category and recipient. Global Health Observatory [online database]. https://www.who.int/observatories/global-observatory-on-health-research-and-development/monitoring/investments-on-grants-for-biomedical-research-by-funder-type-of-grant-health-category-and-recipient (accessed 28 March 2026).

[89] Mughal NA, Hussain MH, Ahmed KS, Waheed MT, Munir MM, Diehl TM, Zafar SN. Barriers to Surgical Outcomes Research in Low- and Middle-Income Countries: A Scoping Review. J Surg Res. 2023 Oct;290:188-196. doi: 10.1016/j.jss.2023.04.017. Epub 2023 Jun 1. PMID: 37269802.

[90] Ellens T. Canada's health data is flowing abroad while Ottawa stalls on AI rules. Policy Options. 2025 Oct 24. https://policyoptions.irpp.org/2025/10/health-data-sovereignty/ (accessed 5 March 2026).

profit rather than co-invest in supportive infrastructures,[91] health/digital literacy, and ethical governance in these regions to promote benefit-sharing. These practices have justice implications, as they may exacerbate existing regional inequalities, foster dependency, and raise concerns of worsening 'digital colonialism.'[92] They highlight the importance of establishing strategies to strengthen LMICs' capacity to collect, analyse, and use data for locally defined priorities. These practices also raise the prospect of counter-practices by communities and nations to challenge and resist certain agendas, such as data commodification, and lead to questions about what forms such counter-practices might take.

As actors that amass data for health research gradually shift from governments, health systems, and research institutions to corporations, this new social order may affect how health research agendas are set, and how researchers and health systems may acquire data (e.g., buy data rather than collect data on their own). It raises questions of who should be included in setting agendas regarding how digital footprints should be used and for what research purpose, what means or frameworks are most ethically appropriate for determining these priorities and policies/agendas (e.g. intellectual property regimes versus ethical and professional standards), and how the monetary, scientific, and social values of digital footprints should be determined. In particular, these new data practices raise questions about how to involve communities and groups of persons in the co-production of research to promote equitable and ethically legitimate research. Recognising the commercial realities of data research may also help us to determine what benefit-sharing approaches that go beyond direct or immediate health benefits may be ethically appropriate in the shifting data-focused health research environment.[93]

- How may the proliferation of digital footprints, combined with the power imbalance shaped by colonial legacies and capitalist logics, present a risk for the unethical and inequitable utilisation of data?[94] What strategies can be used to promote equitable use of digital footprint data in health research?

- How may decolonial and strength-based lenses provide foundations for new governance structures or other opportunities that can help resist digital health coloniality and promote health justice?[95]

---

[91] Sassmannshausen F. 2026. As WHO Debates Global AI Regulation, States Clash Over 'Data Sovereignty'. https://healthpolicy-watch.news/who-debates-global-ai-rules/ (accessed 5 March 2026).
[92] Ndemo B. Addressing digital colonialism: A path to equitable data governance. Unesco Policy Lab. Aug 8 2024. https://community.unesco.org/inclusivepolicylab/s/thinkpiece/addressing-digital-colonialism-a-path-to-equitable-data-governance-MCILJ2OCE56JB5ZCY2ZY4M7AZUYE.
[93] Haddow G, Laurie G, Cunningham-Burley S, Hunter KG. Tackling community concerns about commercialisation and genetic research: a modest interdisciplinary proposal. Soc Sci Med. 2007 Jan;64(2):272-82. doi: 10.1016/j.socscimed.2006.08.028. Epub 2006 Oct 13. PMID: 17050056.
[94] Sekalala S, Chatikobo T. Colonialism in the new digital health agenda. BMJ Global Health. 2024;9:e014131. https://doi.org/10.1136/bmjgh-2023-014131.
[95] Sekalala S, Chatikobo T. Colonialism in the new digital health agenda. BMJ Global Health. 2024;9:e014131. https://doi.org/10.1136/bmjgh-2023-014131.

- What data practices and governance structures in global health partnerships would best promote shared decision-making, fair benefit-sharing, and respect for community sovereignty?

- How can the involvement of corporations in the digital footprints landscape be properly governed to prevent conflict of interest and promote equitable access to data for health research?[96]

- What ethical governance structures may balance legal protections for corporations and technical data (e.g., intellectual property rights) and fair access to benefit-sharing?

3) Consent, privacy, and datafication

These broader ethical issues arising from the use of multi-sector digital footprints raise questions regarding whether existing research ethics approaches are adequate in navigating these new data practices. One of the key ethical principles for biomedical and health-related research involving humans is respect for persons, which requires that people who are capable of deliberation about their personal choices should be treated with respect for their capacity for self-determination. For individuals with impaired or diminished autonomy, or who may be dependent or experience various forms of vulnerability, respect for persons requires that they are afforded security against harm or abuse.[97] The use of digital footprints in health research in LMICs requires revisiting established informed consent and privacy protection practices that have been core operationalisations of these principles.

In health research with identifiable participants, respect for persons and autonomy is typically operationalised through voluntary informed consent, which involves providing prospective participants clear information about the purpose of the research, study procedures, how their data will be collected, used, shared, and retained, and any potential risks of participation or withdrawal. Participants must be free to decline involvement in any aspect of the research without negative consequences. In longitudinal studies, consent is often obtained for different components or at multiple time points, recognising that participants may agree to some uses of their data but not others, such as sharing data beyond the primary research team. Where feasible, participants should also be able to request removal of their data if they withdraw.

In the context of secondary use of digital footprints in health research, questions abound how consent should be ethically managed when data originally collected for other purposes are later used for health research. While broad informed consent for

[96] Sekalala, S., Rawson, B., & Andanda, P. A socio-legal critique of the commercialization of digital health in Sub-Saharan Africa. Policy Studies. 2026 47(2), 285–305. https://doi.org/10.1080/01442872.2025.2451966.

[97] International Ethical Guidelines for Health-related Research Involving Humans, Fourth Edition. Geneva. Council for International Organizations of Medical Sciences (CIOMS); 2016. https://cioms.ch/publications/product/international-ethical-guidelines-for-health-related-research-involving-humans/.

unspecified future use may be feasible for data generated in the formal health system or health research settings, it is more challenging in the context of other digital footprints. Individuals who use wearable devices, AI prompts, online search engines, or social media for personal purposes may have little awareness that their data may be repurposed for health research. These dynamics raise important questions regarding how key actors can best respect people and demonstrate system trustworthiness in the context of using digital footprints,[98] when individual consent and absolute anonymity are simply no longer achievable. Two key concerns include the possibility of re-identification and datafication of health, the latter of which is the process where individuals' activity, behaviour, and experiences are recorded as quantified data and made analysable as reference points for health.[99]

First, even when digital datasets are aggregated and de-identified, linking multiple sources of real-time data can enable the re-identification of individuals. For instance, cellphone metadata that are nominally anonymised can still reveal distinctive geolocation patterns. When combined with device identifiers and social media data, they may disclose highly detailed personal information, such as where individuals live, work, travel, and with whom they interact. When linked with consumer data or social media activities, they may uniquely identify individuals and their personal relationships, creating the possibility that sensitive information about health status, behaviours, or social circumstances could be traced back to them. As the cross-sector ecosystem further heightens these risks, we need to consider what would be ethically acceptable risks, how these potential dangers should be communicated to the public, and how privacy can still be best protected.

Second, intersecting with re-identification risks, datafication of health has important ethical implications. Datafication may allow researchers to infer intimate details about health conditions or risks without individuals necessarily knowing or explicitly disclosing them. Digital footprints may serve as proxies for individuals' social and physical behaviors to infer mental health conditions, sexual or reproductive health concerns, chronic illness, inheritable disorders, or other stigmatising conditions that have social and relational consequences. If high-reliability data and sophisticated analytical models can indeed provide accurate insights, they may offset some privacy risks by offering valuable opportunities for health research that can enhance clinical interventions and health policies. Nonetheless, in the context of unknown quality and speculative utility, whereby digital footprints may become data shadows that have the potential of misrepresenting the subjects via faulty proxies or other methodological pitfalls,[100] there are ethical and epistemological questions about whether individuals have meaningful

---

[98] Harvey K, Laurie G. Proxies of Trustworthiness: A Novel Framework to Support the Performance of Trust in Human Health Research. J Bioeth Inq. 2024 Dec;21(4):625-645. doi: 10.1007/s11673-024-10335-1. Epub 2024 Mar 29. PMID: 38551757; PMCID: PMC11882638.

[99] Ada Lovelace Institute. 2020. The data will see you now: Datafication and the boundaries of health. https://www.adalovelaceinstitute.org/wp-content/uploads/2020/11/The-data-will-see-you-now-Ada-Lovelace-Institute-Oct-2020.pdf (accessed 5 March 2026).

[100] Milne R, Costa A, Brenman N. Digital phenotyping and the (data) shadow of Alzheimer's disease. Big Data Soc. 2022 Jan;9(1):20539517211070748. doi: 10.1177/20539517211070748. Epub 2022 Jan 11. PMID: 36793447; PMCID: PMC7614175.

control over the linkage and interpretation of such sensitive information. If meaningful control, through consent or otherwise, is no longer possible, how can respect still be shown in ways that can justify data uses for research purposes?

These concerns are particularly salient in LMICs, where access to treatment or support services may be limited, such that individuals whose data contribute to research may experience individual and/or familial privacy risks without necessarily benefiting from appropriate health interventions. In increasingly interconnected data ecosystems, where information flows across multiple platforms and institutions with limited transparency, datafication of health may not only threaten privacy but also shape how individuals understand their personal identities, health conditions, and social vulnerabilities.

Take menstruation and pregnancy tracking apps as an example. These technologies can show when users' period or pregnancy stops and starts. We can imagine a study on environmental impacts on menstruation using period tracking data, which may inadvertently reveal that someone has had an abortion or miscarriage. Geolocation data and internet search histories may also suggest that a user is trying to seek care for pregnancy loss or termination. While real-time digital data can help health researchers learn more about environmental and population reproductive health needs, in places where abortion is illegal or restricted, or where miscarriage is socially stigmatised and has familial implications, there are heightened intersecting privacy and prejudice concerns, as the use of consumer footprints in health research may expose people in certain areas to social or legal risk.[101] They raise important ethical questions of whether there are particular justice-related responsibilities owed to socially stigmatised or marginalised populations for whom harms may be greater from the use of digital footprints in health research.

It is worth noting that the framing of consent and privacy described above is primarily anchored in an individualist conception of the person as the primary unit of moral and legal concern, which has shaped the international bioethics architecture from the Declaration of Helsinki to CIOMS. However, many people in the world live by moral traditions in which personhood, responsibility, and social life are understood in fundamentally relational terms (e.g. Muslim, African, indigenous, and other non-Western traditions). The self is embedded in a web of familial, tribal, and communal relationships that constitute part of the person's moral identity and carry their own claims and obligations. Some inferred conditions from digital footprints may lead to stigma, discrimination, legal disputes, and disruption of family structures that extend far beyond the individual whose data are collected and used. Data about people other than the primary monitored users (e.g., family caregivers) may also be captured by ambient sensing technologies, raising relational justice questions.[102] Decisions about one's own data are thus not always purely individual decisions: they affect and implicate others

[101] Torchinsky R. How period tracking apps and data privacy fit into a post-Roe v. Wade climate. NPR. 2022 May 10 (updated 2022 Jun 24). https://www.npr.org/2022/05/10/1097482967/roe-v-wade-supreme-court-abortion-period-apps (accessed 5 March 2026).
[102] Ho A. A Relational Justice Approach for Ambient Intelligence for Elder Care. Am J Bioeth. 2026 Feb;26(2):22-24. doi: 10.1080/15265161.2025.2608626. Epub 2026 Feb 12. PMID: 41678683.

who share one's family name, tribal affiliation, or community standing.[103] This raises the question: What would a governance framework adequate to this reality look like? At a minimum, research ethics review of health research using digital footprint data should systematically ask: does this research generate data with a wide 'moral sensitivity radius'?[104] If so, what mechanisms beyond individual consent are required? These might include family or community consultation processes, restrictions on secondary use of high-sensitivity data categories, and culturally-informed guidance to research ethics committees on when individual consent is insufficient.

- Are traditional concepts like consent and privacy still fit for purposes in the digital footprints context? If so, how so? If not, how should we respond to the ethical concerns about respect that they are designed to address?

- When health systems and researchers are using not only aggregate metadata but also individual level digital footprint data, what are the implications for privacy and risks surrounding sensitive health information?

- How should we balance the potential public interests of data-sharing and the potential risks of disclosure of private and sensitive information about users?

- Given evolving data gathering and sharing practices, how are privacy and consent norms changing in the context of expanding availability of global digital footprints?

4) Ethical governance

The aforementioned ethical considerations lead to the overarching question regarding how the use of digital footprints should be governed. Historically, ethical governance of research involving humans and associated data practices has relied on international and local ethical guidelines for biomedical research, local laws and regulations, and on healthcare or academic institutional review mechanisms. Nonetheless, the rapid expansion of big data infrastructures and AI analytics is transforming how digital footprints are collected, linked, analysed, and shared across institutional and sectoral boundaries for health research purposes far removed from the context in which it was originally collected. As public-private initiatives continue to grow in introducing non-health data into the health research space, there are questions of what ethical responsibilities commercial entities may have in processing personal data,[105] and what "good governance" should look like to address intersecting ethical issues that arise from using digital footprints, as ethical oversight may span multiple sectors populated by

[103] For a parallel discussion on these issues in the genomic context, please see Ghaly, M. *Islamic Ethics and Incidental Findings*: *Genomic Morality Beyond the Secular Paradigm*. Springer Nature. 2024.
[104] The concept of a 'moral sensitivity radius' invites reflection on the breadth of persons, relationships, and social structures that a given category of data can affect beyond the individual from whom it is collected. Different data types carry different moral sensitivity radii, with potential implications for consent, governance, privacy, and harm prevention.
[105] McCoy MS, Allen AL, Kopp K, Mello MM, Patil DJ, Ossorio P, Joffe S, Emanuel EJ. Ethical Responsibilities for Companies That Process Personal Data. Am J Bioeth. 2023 Nov;23(11):11-23. doi: 10.1080/15265161.2023.2209535. Epub 2023 Jun 1. PMID: 37262312.

actors with divergent interests, levels of accountability, and understandings of ethical responsibility. There are questions around 1) access to data, such as who should be able to access data from companies or consumers and on what basis, and 2) accountability for data use, including who is responsible for the use and misuse of digital footprints. These challenges are particularly salient when considering the public interest dimension of data-driven research.[106] Many forms of data-intensive health research rely on large, aggregated datasets capable of generating insights at the population level. Such research may help to improve disease surveillance and health systems response. At the same time, the use of large-scale datasets from digital footprints can create tensions between collective benefits and individual interests, particularly when personal information is repurposed without individual knowledge across multiple studies or combined with data from non-health sectors such as consumer behaviour or digital platforms. There is the perennial quality issue when diverse data sources are used, as biases and inaccuracies may occur not only at the data source,[107] but also during processing due to methodological limitations and pitfalls,[108] and these might also negatively impact metadata. The cross-sectoral nature of research involving digital footprints raises multiple governance challenges that range from 1) how to govern research well across such diverse sectors to 2) how to govern proportionately to avoid overburdening researchers and potentially impeding socially valuable health research.

Ethical governance challenges are heightened within the context of ongoing global inequities and power asymmetries between HICs and LMICs. For example, AI systems may be trained on globally sourced datasets that include information from populations in LMICs, while the resulting technologies, economic value, and scientific recognition accrue primarily to institutions in HICs. Conversely, training might focus only on certain (privileged) data sets, thereby producing research sources and results that are inherently skewed or even biased. This dynamic raises questions of what governance approaches may best attend not only to technological innovation but also to justice, reciprocity, and equitable participation in the cross-sector and global data ecosystem. Good governance is not simply about minimising harm. It is also about enabling open science and open research that is ethical, responsible, and enhances fair distribution of benefits.

At the implementation level, there are also questions of whether or how there may be interoperable governance, i.e., mutually acceptable governance standards that can promote ethical research without replicating cumbersome approval mechanisms that may hinder good research. Miscellaneous data sources and variable levels of identifiability (e.g., identifiable data vs metadata) may warrant proportional governance

---

[106] Laurie, G. and Stevens, L. Developing a Public Interest Mandate for the Governance and Use of Administrative Data in the United Kingdom. Journal of Law and Society. 2016 43: 360-392. https://doi.org/10.1111/j.1467-6478.2016.00759.x.

[107] Ho A. *Live Like Nobody is Watching: Relational Autonomy in the Age of Artificial Intelligence Health Monitoring*. New York: Oxford University Press. 2023.

[108] Olteanu A, Castillo C, Diaz F, Kıcıman E. Social Data: Biases, Methodological Pitfalls, and Ethical Boundaries. Front Big Data. 2019 Jul 11;2:13. doi: 10.3389/fdata.2019.00013. PMID: 33693336; PMCID: PMC7931947.

models that provide robust and efficient means of paying due regard to both privacy and the public interests in research.[109] We welcome case studies and normative papers that tackle this specific issue, and normative papers that explore ways to navigate complexity, deliver protection and support promotion of good research that do not result in unnecessary duplication of effort.

- How may data-driven health research using digital footprints raise ethically relevant concerns that differ from those addressed by traditional biomedical research frameworks, especially in a cross-sectoral context?

- How should ethical governance grapple with balancing individual autonomy, privacy, and data control with broader societal interests in using digital footprints to advance health research that may improve health outcomes?

- What governance or oversight mechanisms would best address the distinctive ethical challenges posed by large-scale data environments?

- As some governments exempt consumer wearables and other health-related devices from research oversight,[110] how may we create the right incentives and governance systems for companies, institutions, and stakeholders to generate, safeguard, and share digital footprints for health research, along with establishing the essential infrastructure and institutions for effective operation?[111]

- What are some examples of best practices in digital data governance that can inform the use of digital footprints in health research in LMICs?

VI. Conclusion

Digital footprints and real-time data present significant opportunities to expand the scope, timeliness, and granularity of health research, particularly in LMICs where traditional data infrastructures may be limited. At the same time, we have seen how their use introduces complex ethical issues around data quality, bias, equity, consent, privacy, datafication, and governance. These interconnected challenges are shaped by the multi-sectoral nature of digital data ecosystems and by enduring global asymmetries in resources, power, and technological capacity.

As digital data practices continue to evolve, there is a pressing need to move beyond existing research ethics models and explore how ethical governance can be effectively developed and implemented across diverse actors and sectors. Using digital footprints to enhance ethical health research in LMICs will require sustained interdisciplinary

---

[109] Sethi N, Laurie GT. Delivering proportionate governance in the era of eHealth: Making linkage and privacy work together. Med Law Int. 2013 Jun;13(2-3):168-204. doi: 10.1177/0968533213508974. PMID: 24634569; PMCID: PMC3952593.

[110] Reuter E. FDA exempts more wearable, AI features from oversight. MedTechDive. 2026 Jan 8. https://www.medtechdive.com/news/fda-exempts-wearable-ai-features-guidance/809099/.

[111] Ndemo B. Addressing digital colonialism: A path to equitable data governance. Unesco Inclusive Policy Lab. 2024 Aug 8. https://community.unesco.org/inclusivepolicylab/s/thinkpiece/addressing-digital-colonialism-a-path-to-equitable-data-governance-MCILJ2OCE56JB5ZCY2ZY4M7AZUYE.

engagement and the development of context-sensitive governance frameworks that can respond to rapid technological change while promoting trust, justice, and global health equity. We look forward to receiving presentation proposals that can help address these issues in the evolving data ecosystem.

16 April 2026