Journal of the American Medical Informatics Association, 00(0), 2021, 1–7
doi: 10.1093/jamia/ocab012
Perspective

Perspective

Digital phenotyping and sensitive health data: Implications for data governance

Ignacio Perez-Pozuelo,[1,2] Dimitris Spathis,[3] Jordan Gifford-Moore,[4] Jessica Morley,[5,6] and Josh Cowls[2,5]

[1]MRC Epidemiology Unit, School of Clinical Medicine, University of Cambridge, Cambridge, United Kingdom, [2]The Alan Turing Institute, London, United Kingdom, [3]Department of Computer Science and Technology, University of Cambridge, Cambridge, United Kingdom, [4]School of Law, University of Oxford, Oxford, United Kingdom, [5]Oxford Internet Institute, University of Oxford, Oxford, United Kingdom, and [6]Nuffield Department of Primary Care, University of Oxford, Oxford, United Kingdom

Corresponding Author: Ignacio Perez-Pozuelo, MRC Epidemiology Unit, School of Clinical Medicine, University of Cambridge, UK, Cambridge CB22 0QQ, UK (Ip325@cam.ac.uk)

Received 5 November 2020; Revised 14 January 2021; Editorial Decision 15 January 2021; Accepted 21 January 2020

Mobile and wearable devices, such as smartwatches and fitness trackers, increasingly enable the continuous collection of physiological and behavioral data that permit inferences about users' physical and mental health. Growing consumer adoption of these technologies has reduced the cost of generating clinically meaningful data. This can help reduce medical research costs and aid large-scale studies. However, the collection, processing, and storage of data comes with significant ethical, security, and data governance considerations. A complex ecosystem is developing, with the need for collaboration among researchers, healthcare providers, and a broad range of entities across public and private sectors, some of which are not traditionally associated with health care. This has raised important questions in the literature regarding the role of the individual as a patient, customer, research participant, researcher, and user when consenting to data processing in this ecosystem.[1] Here, we use the emerging concept of "digital phenotyping"[2] to highlight key lessons for data governance that draw on parallels with the history of genomics research, while highlighting areas in which digital phenotyping will require novel governance frameworks.

UBIQUITOUS PERSONAL HEALTH DATA

Phenotypic traits are broadly defined as the observable characteristics of an individual that arise from the combined effects of their genotype and the environment. Analysis of phenotypes yields important insights across many fields of research, including human evolution and cultural history, the identification of the genetic basis of disease and health-related traits, drug repurposing, and pharmacogenomics. Building on developments made through the collection and analysis of extended phenotypic data through the growth and evolution of digital products,[3] digital phenotyping can be defined as the "moment-by-moment quantification of the individual-level human phenotype using data from personal digital devices."[2,4] This process is often passive and allows for the quantification of the individual-level behavioral phenotype through personal digital devices such as mobile phones and wearable technologies.[4] Advances in these data collection tools have accelerated across both academia and industry, along with diverse applications in clinical and public health settings. While passive data generated and collected through mobile or wearable devices are not without limitations, for instance, with regard to causality or the ability to match its outcomes to that of clinical outcomes or diagnoses, its use can overcome some of the issues associated with traditional survey-based methods. For instance, the ability to obtain in situ data offers significant opportunities to mitigate the well-documented issues of self-reporting inaccuracies,[5] inconsistent classification and recording of phenotypic data,[6] and behavior modification in some contexts due to participation in an observed environment.[7] Further, these tools enable, at an unprecedented scale, long-term phenotyping in free-living conditions with the potential for reduced subject attrition.[8]

Early examples of digital phenotyping studies include large-scale involuntary hand tremor analysis via mouse cursor movement[9] and the use of Microsoft Bing search queries to detect neurodegenerative conditions.[10] Despite improvements in the collection and classification of data, digital phenotyping poses its own unique risks for users. Given the multidisciplinary nature of the field and the differ-

ent levels of sensitivity of the data being collected, digital phenotyping interacts with a broad range of laws and governance regimes, ranging from medical and research ethics to contract law and data protection regulation. The international adoption of consumer devices allowing digital phenotyping research adds additional complexity in determining the applicable regulatory framework for data collection, sharing, and analysis. Although there are established processes in medical research for international data sharing, the longitudinal and dynamic nature of digital phenotyping can itself create ongoing obligations across different jurisdictions and spheres of regulation. As such, there is a risk that consumers will be insufficiently protected if they are exposed to digital phenotyping technologies that do not fall neatly within any existing consumer protection regime with an effective enforcement framework.

As devices enabling digital phenotyping research are often consumer electronics, in scenarios outside of institutional research frameworks an important dichotomy arises between consumers' motivation to use these technologies and technology providers' incentives to collect, analyze, share, or monetize data produced by users.[11] Advances in consumer electronics sensors and trends in wellness technologies has brought health and lifestyle data that might traditionally have been governed by medical research ethics and regulations outside of these institutional settings. Data that permit inferences about users' health or lifestyle through digital phenotyping are now increasingly available for collection by commercial hardware and software vendors, which are not typically healthcare providers. A broad range of harms related to the collection of health data online has been highlighted in the academic and policy literature, including unethical data collection[12] and provision of inaccurate clinically relevant data.[13] As digital phenotyping becomes more prevalent and is used by commercial providers of other services, or to generate diagnoses, there is also the potential for discriminatory use of sensitive data, such as exclusionary insurance, employment discrimination, or unfair credit scoring.[14,15] While a number of existing data protection, consumer protection, and antidiscrimination laws may help safeguard the use of personal health data in various contexts, the efficacy of these laws have not been fully tested in the array of novel contexts in which health data may be used. Similarly, a lack of regulatory clarity and oversight may also fail to provide commercial providers and researchers with the certainty required for ethical scientific research and innovation and shift the burden of screening digital health technologies onto patients and clinicians.[16] Further, digital phenotyping must move toward the development of standards that identify sources of bias and enforce the development of models that are robust to skews and incompleteness associated to this type of data.[17,18]

Digital phenotyping at scale

While genotyping has become more widely accessible during the past 20 years due to falling costs of sequencing and consumer-focused providers, many of the benefits of digital phenotyping arise from technologies developed first for mass consumer adoption. From a regulatory perspective, one difficulty posed by digital phenotyping is the use of data collected outside of a traditional healthcare context to make health- and wellness-related inferences. This expands the circumstances in which health data are collected, and also importantly allows for health-related analysis of types of data that may previously have been considered less sensitive. While academic research using consumer technologies remains subject to existing frameworks on ethical research, the increasing functionality of consumer electronics means that digital phenotyping can also be conducted in a commercial context. There is already some scope for supervision by consumer protection agencies with relatively broad remit, such as data protection authorities, in Europe under the General Data Protection Regulation (GDPR), and to an extent the Federal Trade Commission (FTC) in the United States, for deceptive or unfair practices regarding data governance.[19] Omnibus data protection regulation such as the GDPR provides a baseline level of protections for data processed outside of a healthcare setting. This is essential for securing data rights in circumstances in which consumer devices are actively designed or can be repurposed to engage in digital phenotyping. However, frameworks such as the U.S. Health Insurance Portability and Accountability Act (HIPAA) place greater emphasis on the context of processing and the parties involved,[20,21] with consumers relying either on more general state law such as the California Consumer Privacy Act (CCPA) or the FTC as a backstop to pursue data governance rights outside of a HIPAA context. As U.S. policymakers react to the growth of digital phenotyping and there are state and federal attempts to introduce broader regulatory frameworks akin to the CCPA, there will be opportunities to learn from the EU experience with the GDPR. Where digital phenotyping allows health-related insights to be drawn from an increasingly diverse range of data, greater regulatory clarity on the categorization and treatment of digital phenotyping data in different contexts, for example, on the scope of EU GDPR definition of personal data concerning health, would allow agencies such as data protection authorities to better allocate scarce resources. While the FTC in the United States has continued to emphasize its supervision of health-related data, even when not covered by HIPAA,[22] there remains no general concept of sensitive data in U.S. general law to provide ex ante guidance to consumers of their rights.[23] For U.S. policymakers, this creates the opportunity to improve on the GDPR in this respect and provide greater clarity around the use of data for digital phenotyping. While governing digital phenotyping at scale may require new models of resource allocation and oversight, initial steps could focus on developing enforceable industry standards, such as approved GDPR codes of conduct, to act as certifications of specific data governance standards for consumers. Given the importance of international collaboration in medical research, it will be essential for policymakers across jurisdictions to consider the obstacles reported by medical researchers in complying with substantially different local data regulations.[24]

Although reforms to existing regulatory frameworks are required, there also remains underenforcement of regulations which are already in force. Even where digital phenotyping is covered by data protection regulations that apply outside of institutional research studies, given the potential scale of the field due to widespread adoption of consumer electronics, regulators lack the resources required to provide comprehensive oversight.[19] A 2019 study found that numerous mobile health apps still regularly failed to disclose processing of special category health data under the GDPR, instead providing only the more basic protections required for less sensitive data.[25] Even among prominent apps more prone to regulatory scrutiny, the complexity of terms and applicable regulations can prevent consumers from understanding the nature of their data being processed. For example, the Fitbit Privacy Policy treats the activity and fitness data that it directly collects as if they were nonhealth data, while noting that for any health data obtained from other sources, or other special category data under the GDPR, Fitbit will notify users and request separate explicit consent to process that data.[26] However, the same Privacy Policy separately informs users

of the possibility that a broad range of collected data, including exercise, activity, sleep, biometric, geolocation, and personally identifying information, may be collected. The result is that data subjects may be unclear when accepting the Privacy Policy what forms of data Fitbit classifies as "health data" at that time and must trust that Fitbit will seek additional explicit consent to process this type of data. While participants in certain forms of institutional research or employer-sponsored programs may benefit from Fitbit, or equivalent hardware providers' research conduct guidelines or HIPAA-compliant offerings, these are unavailable to consumers using the same device outside of the settings contemplated by institutional research or HIPAA. While legislation such as the CCPA provides more explicit guidance on which forms of data are subject to certain rights and goes beyond the FTC's narrower remit, the CCPA itself has no dedicated enforcement agency that can issue engage with issues of interpretation or conduct investigations. It has been argued that this sector-specific approach without a designated regulatory contact point created obstacles to rapid U.S. public–private sector collaboration in responding to coronavirus disease 2019 (COVID-19) in 2020.[27]

Moreover, despite initially being developed as a consumer device, Fitbit and similar wearable devices are increasingly used to generate health-related insights in both research and consumer settings. However, there are few agreed standards and minimal regulatory oversight over the necessary reliability of outputs for research and clinical purposes. Although classification as a "medical device" introduces requirements regarding the validity of results, longitudinal reporting, notification to users of serious health concerns, and improved reporting, only some device manufacturers have elected to seek classification as a medical device (eg, Apple Watch's electrocardiogram app obtained de novo Food and Drug Administration clearance in the United States, and is classified as a class II medical device), and often only for some device functions.[28] The methodology for classification as a medical device differs across jurisdictions, and the distinction between a medical device and a "wellness device" can depend on the manufacturer's intended uses for the device.[29] Particularly for software products, manufacturers can encounter difficulties navigating complex regulation, and products that require classification as a medical device may still be available for public access without appropriate oversight.[30] As a result, consumers may be under the impression that a product has been subject to a greater degree of regulatory scrutiny with regard to the quality of its data measurement and analysis than is necessarily the case.

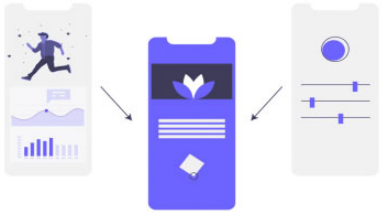
The Fitbit Research Pledge is an example of a consumer device provider explicitly applying aspects of institutional research frameworks to the use of their device, but it is not yet clear that new binding obligations are imposed or that this applies outside of formal studies and published research. Although the FTC has broad jurisdiction over similar forms of representations to consumers,[19] the Research Pledge appears to apply only in institutional research settings where those rights would already be provided as part of the institutional review board process. Particularly when outside of an institutional research setting, digital phenotyping is vulnerable to many of the common problems in mobile health. Companies collecting sensitive health data regularly make unilateral changes to their terms of service, and privacy disclosures are frequently inadequate, underscoring a lack of protection for personal data and user privacy.[31] Undisclosed sharing of digital phenotyping data, including linkable identifiers, is prevalent.[25] Inconsistent regulatory oversight, unclear terms and conditions, and failure to disclose data sharing and secondary use can limit the ability for healthcare professionals to recommend otherwise beneficial apps in fields such as mental health care.[25] In the research context, the GDPR adopts lower protections for data subjects in which the purpose of processing is solely for statistical or scientific research purposes. Despite submissions from some concerned groups, such as the BioMolecular Resources Research Infrastructure–European Research Infrastructure Consortium regarding the need to define "scientific research" to exclude some forms of commercial processing,[32] the GDPR was passed to also allow commercial providers to use this research exemption to process sensitive personal data without consent (though still subject to EU Member State law, technical safeguards, and research ethical standards). Clear policy guidance on data sharing practices in these instances is critical to maintaining public trust in scientific digital phenotyping research and enabling the use of these methods for clinical care.

THE RISKS OF DIGITAL PHENOTYPING: LESSONS FROM GENETICS

When considering improvements to the framework for digital phenotyping, there are also valuable precedents from an earlier wave of health technology innovation. Advances in genotyping techniques, particularly from the 1990s onward, created an extraordinary opportunity to better understand human health. At the same time, the sharing of the sensitive individual-level health data required for scientific advances created the need to develop new standards, policies, and regulations for genetics and bioinformatics research. This allowed policymakers in some jurisdictions to enact measures such as obligatory genetic counseling, requirements for validity of results, informed consent, and chain of custody procedures,[33] which built on iterative resources such as the Bermuda Principles,[34] Oviedo Convention,[35] genetic testing protocols,[36] and Council for International Organizations of Medical Sciences guidelines.[37]

To advance human health and infectious disease research, cross-border data sharing has become essential in genomics, leading to the creation of a variety of genomic data resources. These databases are mainly constructed by and for publicly funded scientific and medical researchers and their institutions. They range from being completely open, like the BRCA exchange, ClinVar, and Genome Aggregation database,[38–40] to having regulated access like the European Genome-Phenome Archive, the dbGaP (database of Genotypes and Phenotypes), and the Human Gene Mutation database.[41–43] Potentially instructive models to draw on for digital phenotyping data include controlled- or managed-access models, data access committees, data safe havens, dynamic and tiered consent, differential access, explicit open-access consent, and portable legal consent. In particular, dynamic and tiered consent models are readily applicable to areas of digital phenotyping, in which the sensors for data collection tend to be associated with a consumer device, such as a mobile phone, which could more easily support a user-friendly interface for dynamic consent models.[44] Through collaboration across researchers, commercial providers, and regulators, it may be possible to leverage these technology platforms to further improve the delivery and application of data management solutions developed in genomics.

A series of studies have demonstrated the challenges for researchers of fully anonymizing data (including in controlled-access databases such as the dbGaP), observing data subjects' bounded consent on collected data, and delivering clinically valid and meaningful data in a direct-to-consumer setting.[45] While the lessons learned



GENOTYPING DATA	DIGITAL PHENOTYPING DATA
Single dataset	Complex, multi-layered/modal datasets
Collected once	Collection ongoing
Unchangeable/immutable	Changeable/mutable
Often knowledge of collection	Sometimes knowledge of collection
Non-negligible cost/expertise	Negligible cost/expertise
Marginal benefits from triangulation	Substantial benefits from triangulation

Figure 1. Differences between genotyping data and digital phenotyping data. Digital phenotyping can never be said to be complete, because new data are generated continuously to reflect changing patterns of user behavior. Although sophisticated data analysis often requires considerable infrastructure and expertise, the cost of processing and analyzing each additional data point is usually negligible.

from genomics in these areas can assist with approaching digital phenotyping data governance, we must also be mindful of the important differences between genotypic and phenotypic data. While genotypic data solely comprise genetic code, digital phenotyping data are extremely diverse. As a result, the data collected under the umbrella of digital phenotyping may give rise to a broader range of possible harms.

Though there are a number of differences between genotyping and digital phenotyping data (see Figure 1), the disparity is especially apparent in the manner of collection. The negligible costs of most types of digital phenotyping data collection after initial investments in infrastructure means that it is efficient to aggregate large datasets from different users who may be unaware of their inclusion in a dataset where personal devices upload data by default to a centralized data controller. Even in cases in which explicit agreement is sought in terms of service for data collection, the terms may not reflect the nature of consent for research or secondary use, and complex terms and conditions can result in user fatigue and a "tick-box" approach, meaning that users are less likely to provide truly informed consent.[31] Moreover, the vast majority of digital phenotyping data arise from commercial products, in which the role of these data and the associated research is at least in part to support a business model. Most of these data are therefore not used to produce pure public goods or knowledge and are not freely available under existing governance frameworks for proprietary data.

The current fragmented approach to regulatory oversight, classification of data for the purpose of identifying the applicable laws, and varying data governance practices lowers user trust in digital phenotyping and limits potential medical research. While the precedents of considered regulation and multistakeholder collaboration in genomics should inform developments in this field, it is also important to improve on these models where possible, and address aspects of digital phenotyping that require novel solutions.

While genetics databases have generally tended toward releasing aggregate data, the unique collection and delivery platforms of digital phenotyping may offer new models for data management and informed research participation (see Table 1). For instance, these technologies include the means to reduce the current practice of centralized data consolidation for the purposes of extracting value. Through privacy-preserving, decentralized methods like federated learning[46–48] and zero-knowledge proofs, users could maintain sole custody of their data. These methods also enable model sharing,[49] as opposed to data sharing, which could allow for more seamless cooperation between corporations and academic or public sector institutions. Advances in general techniques that are applicable beyond digital phenotyping such as differential privacy techniques could also address this issue by collecting and aggregating information about groups of users' habits and behaviors while not sharing data from individual users. Similarly, from a consumer-facing perspective, digital phenotyping technologies could enable innovation in dynamic consent and through modern user interfaces and devices. These techniques remain the subject of ongoing academic research and improvement in their application to digital phenotyping. For example, differential privacy techniques have been found to be difficult to apply in practice by some healthcare researchers[50] and dynamic consent models can lead to underproduction of data for secondary studies.[51] These trade-offs demonstrate the complex challenges posed by the field and indicate that regulatory frameworks ought to clearly prescribe the forms of digital phenotyping data to which they apply, while remaining principles-based and technology neutral in their requirements for data governance.

The COVID-19 pandemic has further complicated the ethics and governance of the collection and use of digital phenotyping data.[52] For example, in some jurisdictions, anyone who carries a smartphone can now be considered a potential transmitter of COVID whose location and contacts with other smartphone users can be traced. This rapid expansion of the types of data that can be considered "health-related" may become entrenched after the pandemic, as forms of behavior that were previously seen as unrelated to health, such as an individual's movement through physical space and their contact with

Table 1. Building on developments in genetics to establish a path for digital phenotyping

	Lessons From Genetics	Digital Phenotyping
Data governance	• Open, tiered, and managed data access • Innovative consent models	• Federated learning and zero-knowledge proofs could allow secure data sharing even between competing organizations • Differential privacy (adding noise to individual datapoints) can be applied, especially where location/GPS data is recorded
Informed consent	• Culture of research ethics and oversight • Clear classification of consent to participation in secondary data use and explanation of the implications	• Transparent presentation of terms, employing modern UX practices • Innovation in dynamic consent and user interfaces for research
Research methods	• Developing standards for the reliability of results in clinical settings • Leveraging data-sharing practices for genome-wide association studies	• FACT AI and models that actively adjust for demographic parity • Provision of reliable outputs and contextualizing clinically relevant information
Supervision and governance models	• Multistakeholder input into applicable guidelines and regulatory frameworks • Scope for public–private collaboration and research exemptions included in applicable laws	• Enabling regulatory oversight through allocation of resources and cross-domain expertise to existing consumer protection regulators • Developing practical industry codes, guidance, and oversight mechanisms (eg, an enforceable GDPR code)
Ongoing challenges	• Direct-to-consumer models and collaboration among clinicians, commercial providers, and researchers • Anonymization of individual-level data • Biobanks that balance the values and rights of participants while ensuring their long-term sustainability.	• Diverse potential harms from data collection and processing due to heterogeneous data • Heightening the differences and disparities of historically marginalized groups • Transparency regarding the validity of inferences and associated behavioral interventions • Commercially led approach to traditionally public research

AI: artificial intelligence; FACT AI: Fairness, Accountability and Transparency in AI; GDPR: General Data Protection Regulation; UX: user experience.

others, are increasingly considered to be relevant for analysis of public health crises such as infectious disease spread.

Given the nature of the data collected, laying the foundations for responsible data governance and providing reliable, well-validated, and contextualized outputs will be critical to building trust and enabling the development of the digital phenotyping field. Mobile and wearable technologies have the potential to transform healthcare[53] by providing low-cost, objective measurements of physical, cognitive, emotional, and social behaviors at unprecedented scale.[54] Nonetheless, several limitations must be overcome if this potential is to be realized, particularly as the development and deployment of digital phenotyping technologies for mobile health has vastly outpaced that of the methodology to evaluate its validity and safeguard users' rights. Several complex issues must first be resolved, such as around who owns, controls, and can use personal health data to derive wider insights; the formats and standards that should underpin how these data are shared; and how the range of potential uses of personal data are explained and justified to data subjects.

FUNDING

IP-P is supported by GlaxoSmithKline and Engineering and Physical Sciences Research Council through an iCase fellowship (17100053). DS is supported by the Embiricos Trust Scholarship of Jesus College Cambridge and the Engineering and Physical Sciences Research Council through grant DTP (EP/N509620/1). JC is the recipient of a doctoral scholarship from The Alan Turing Institute. JM is supported by the Wellcome Trust.

AUTHOR CONTRIBUTIONS

IP-P, DS and JC developed the initial concept of the article. IP-P led the manuscript writing and finalized the article. JM and JC led the digital ethics discussion, IP-P and DS led the technical considerations, and JGM led the legal scholarship, while all authors provided with edits to the contents of this manuscript. IP-P and DS created the figure and table with inputs from all other authors. The final manuscript was approved by all authors.

CONFLICT OF INTEREST STATEMENT

The authors declare no competing interests.

DATA AVAILABILITY

No new data were generated or analyzed in support of this research.

REFERENCES

1. Dove E, Chen J. Should consent for data processing be privileged in health research? A comparative legal analysis. *Int Data Privacy Law* 2020; 10 (2): 117–31.

2. Onnela JP, Rauch SL. Harnessing smartphone-based digital phenotyping to enhance behavioral and mental health. *Neuropsychopharmacology* 2016; 41 (7): 1691–6.
3. Jain SH, Powers BW, Hawkins JB, Brownstein JS. The digital phenotype. *Nat Biotechnol* 2015; 33 (5): 462–3.
4. Torous J, Kiang MV, Lorme J, Onnela JP. New tools for new research in psychiatry: a scalable and customizable platform to empower data driven smartphone research. *JMIR Ment Health* 2016; 3 (2): e16.
5. Bauhoff S. Systematic self-report bias in health data: impact on estimating cross-sectional and treatment effects. *Health Serv Outcomes Res Method* 2011; 11 (1–2): 44–53.
6. Girdea M, Dumitriu S, Fiume M, *et al.* Pheno tips: patient phenotyping software for clinical and research use. *Hum Mutat* 2013; 34 (8): 1057–65.
7. McCambridge J, Witton J, Elbourne DR. Systematic review of the Hawthorne effect: new concepts are needed to study research participation effects. *J Clin Epidemiol* 2014; 67 (3): 267–77.
8. Onnela JP. Opportunities and challenges in the collection and analysis of digital phenotyping data. *Neuropsychopharmacology* 2021; 46: 45–54.
9. White RW, Horvitz E. Population-scale hand tremor analysis via anonymized mouse cursor signals. *NPJ Digit Med* 2019; 2 (1): 93.
10. White RW, Doraiswamy PM, Horvitz E. Detecting neurodegenerative disorders from web search signals. *NPJ Digit Med* 2018; 1 (1): 8.
11. Pasquale F. Grand bargains for big data: The emerging law of health information. *MD L Rev* 2012; 72: 682.
12. Troiano A. Wearables and personal health data: putting a premium on your privacy. *Brooklyn Law Rev* 2017; 82 (4): 6.
13. Wang R, Blackburn G, Desai M, *et al.* Accuracy of wrist-worn heart rate monitors. *JAMA Cardiol* 2017; 2 (1): 104–6.
14. Ajunwa I. Algorithms at work: productivity monitoring applications and wearable technology as the new data-centric research agenda for employment and labor law. *Louis ULJ* 2018; 63: 21.
15. Montgomery K, Chester J, Kopp K. Health wearables: ensuring fairness, preventing discrimination, and promoting equity in an emerging Internet-of-Things environment. *J Information Policy* 2018; 8: 34–77.
16. Mathews SC, McShea MJ, Hanley CL, Ravitz A, Labrique AB, Cohen AB. Digital health: a path to validation. *NPJ Digit Med* 2019; 2 (1): 38.
17. Zou J, Schiebinger L. (2018). AI can be sexist and racist—it's time to make it fair. *Nature* 2018; 559: 324–6.
18. Mitchell M, Wu S, Zaldivar A, *et al.* Model cards for model reporting. In: *FAT19: Proceedings of the Conference on Fairness, Accountability, and Transparency*; 2019: 220–9.
19. Solove DJ, Hartzog W. The FTC and the new common law of privacy. *Colum L Rev* 2014; 114: 583.
20. 45 C.F.R. § 160.103. https://www.govinfo.gov/content/pkg/CFR-2014-title45-vol1/pdf/CFR-2014-title45-vol1-sec160-103.pdf Accessed October 18, 2020.
21. See, eg, U.S. Department of Health & Human Services, 'Health App Use Scenarios & HIPAA' (February 2016) https://www.hhs.gov/sites/default/files/ocr-health-app-developer-scenarios-2-2016.pdf.
22. Rich JL. Prepared Statement of the Federal Trade Commission on Opportunities and Challenges in Advancing Health Information Technology. 2016. https://www.ftc.gov/system/files/documents/public_statements/941063/160322commtestimonyhealthinfo.pdf. Accessed October 18, 2020.
23. Schwartz PM, Solove DJ. Reconciling personal information in the United States and European Union. *Calif L Rev* 2014; 102: 877.
24. Bovenberg J, Peloquin D, Bierer B, Barnes M, Knoppers BM. How to fix the GDPR's frustration of global biomedical research. *Science* 2020; 370 (6512): 40–2.
25. Huckvale K, Torous J, Larsen ME. Assessment of the data sharing and privacy practices of smartphone apps for depression and smoking cessation. *JAMA Netw Open* 2019; 2 (4): e192542.
26. Fitbit Legal Privacy Policy. Fitbit.com. 2020. https://www.fitbit.com/us/legal/privacy-policy. Accessed March 20, 2020.
27. Bradford LR, Aboy M, Liddell K. COVID-19 contact tracing apps: a stress test for privacy, the GDPR and data protection regimes. *J Law Biosci* 2020; 7 (1): lsaa034.
28. Piwek L, Ellis DA, Andrews S, Joinson A. The rise of consumer health wearables: promises and barriers. *PLoS Med* 2016; 13 (2): e1001953.
29. Thorogood A, Touré SB, Ordish J, Hall A, Knoppers B. Genetic database software as medical devices. *Hum Mutat* 2018; 39 (11): 1702–12.
30. van Drongelen A, de Bruijn A, Roszek B, Vonk R. Apps under the medical devices legislation. RIVM Letter report 2018-0083. Bilthoven, the Netherlands: The Netherlands National Institute for Public Health and the Environment; 2018.
31. Roberts JL, Hawkins J. When health tech companies change their terms of service. *Science* 2020; 367 (6479): 745–6.
32. Viertler C, Zatloukal K. Biobanking and biomolecular resources research infrastructure (BBMRI). Implications for pathology. *Pathologe* 2008; 29 (S2): 210–3.
33. Kalokairinou L, Howard HC, Slokenberga S, *et al.* Legislation of direct-to-consumer genetic testing in Europe: a fragmented regulatory landscape. *J Community Genet* 2018; 9 (2): 117–32.
34. Borghi M. (2018). Individual rights and property rights in human genetic databases. In: Arnold R, Cippitani R, Colcelli V, eds. Genetic Information and Individual Rights. Regensburg, Germany: Regensburg University Press; 2018: 116–29.
35. Convention for the protection of Human Rights and Dignity of the Human Being with regard to Application of Biology and Medicine: Convention on Human Rights and Biomedicine. ETS No 164. 1997.
36. Lwoff L. Council of Europe adopts protocol on genetic testing for health purposes. *Eur J Hum Genet* 2009; 17 (11): 1374–7.
37. World Health Organization and Council for International Organizations of Medical Sciences. *International Ethical Guidelines for Health-Related Research Involving Humans*. Geneva, Switzerland: WHO Press; 2016.
38. Cline MS, Liao RG, Parsons MT, BRCA Challenge Authors, *et al.* BRCA challenge: BRCA exchange as a global resource for variants in BRCA1 and BRCA2. *PLoS Genet* 2018; 14 (12): e1007752.
39. Landrum MJ, Lee JM, Benson M, *et al.* ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res* 2018; 46 (D1): D1062–D1067.
40. Karczewski K, Francioli L. (2017). The genome aggregation database (gnomAD). MacArthur Lab. https://macarthurlab.org/2017/02/27/the-genome-aggregation-database-gnomad/. Accessed July 17, 2020.
41. Lappalainen I, Almeida-King J, Kumanduri V, *et al.* The European Genome-Phenome archive of human data consented for biomedical research. *Nat Genet* 2015; 47 (7): 692–5.
42. Tryka KA, Hao L, Sturcke A, *et al.* NCBI's database of genotypes and phenotypes: dbGaP. *Nucl Acids Res* 2014; 42 (D1): D975–D979.,
43. Stenson PD, Mort M, Ball EV, Shaw K, Phillips AD, Cooper DN. The Human Gene Mutation Database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. *Hum Genet* 2014; 133 (1): 1–9.
44. Bot BM, Suver C, Neto EC, *et al.* The mPower study, Parkinson disease mobile data collected using ResearchKit. *Sci Data* 2016; 3: 160011.
45. Arias J, Pham-Kanter G, Campbell E. The growth and gaps of genetic data sharing policies in the United States. *J Law Biosci* 2015; 2 (1): 56–68.
46. Yang Q, Liu Y, Chen T, Tong Y. Federated machine learning: Concept and applications. *ACM Trans Intell Syst Technol* 2019; 10 (2): 1–19.
47. Sheller MJ, Edwards B, Reina GA, *et al.* Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data. *Sci Rep* 2020; 10 (1): 12598.
48. Rieke N, Hancox J, Li W, *et al.* The future of digital health with federated learning. *NPJ Digit Med* 2020; 3 (1): 119.
49. Zhang P, Schmidt DC, White J, Lenz G. Blockchain technology use cases in healthcare. *Adv Comput* 2018; 111: 1–41.
50. Dankar FK, El Emam K. Practicing differential privacy in health care: A review. *Trans Data Priv* 2013; 6 (1): 35–67.
51. Wilbanks JT. Electronic informed consent in mobile applications research. *J Law Med Ethics* 2020; 48 (1_suppl): 147–53.

52. Morley J, Cowls J, Taddeo M, Floridi L. Ethical guidelines for COVID-19 tracing apps. *Nature* 2020; 582(7810): 29–31.
53. Perez-Pozuelo I, Spathis D, Clifton EA, Mascolo C. Wearables, smartphones, and artificial intelligence for digital phenotyping and health. In: Syed-Abdul S, Zhu X, Fernandez-Luque L, eds. *Digital Health: Mobile and Wearable Devices for Participatory Health Applications*. Netherlands: Elsevier; 2020: 33–54.
54. Spathis D, Servia-Rodriguez S, Farrahi K, Mascolo C, Rentfrow J. Passive mobile sensing and psychological traits for large scale mood prediction. In: *Proceedings of the 13th EAI International Conference on Pervasive Computing Technologies for Healthcare*; 2019: 272–281; Trento, Italy.