Harmonizing Regulatory Spheres to Overcome Challenges for Governance of Patient-generated Health Data in the Age of Artificial Intelligence and Big Data

Jenifer Sunrise Winter Elizabeth Davidson

University of Hawaii at Manoa

Abstract

Patient-generated health data (PGHD), created and captured from patients via wearable devices and mobile apps, are proliferating outside of clinical settings. Examples include sleep tracking, fitness trackers, continuous glucose monitors, and RFID-enabled implants, with many additional biometric or health surveillance applications in development or envisioned. These data are included in growing stockpiles of personal health data being mined for insight via big data analytics and artificial intelligence/deep learning technologies. Governing these data resources to facilitate patient care and health research while preserving individual privacy and autonomy will be challenging, as PGHD are the least regulated domains of digitalized personal health data (U.S. Department of Health and Human Services, 2018). When patients themselves collect digitalized PGHD using "apps" provided by technology firms, these data fall outside of conventional health data regulation, such as HIPAA. Instead, PGHD are maintained primarily on the information technology infrastructure of vendors, and data are governed under the IT firm's own privacy policies and within the firm's intellectual property rights. Dominant narratives position these highly personal data as valuable resources to transform healthcare, stimulate innovation in medical research, and engage individuals in their health and healthcare. However, ensuring privacy, security, and equity of benefits from PGHD will be challenging. PGHD can be aggregated and, despite putative "deidentification," be linked with other health, economic, and social data for predictive analytics. As large tech companies enter the healthcare sector (e.g., Google Health is partnering with Ascension Health to analyze the PHI of millions of people across 21 U.S. states), the lack of harmonization between regulatory regimes may render existing safeguards to preserve patient privacy and control over their PHI ineffective. While healthcare providers are bound to adhere to health privacy laws, Big Tech comes under more relaxed regulatory regimes that will facilitate monetizing PGHD. We explore three existing data protection regimes relevant to PGHD in the United States that are currently in tension with one another: federal and state health-sector laws, data use and reuse for research and innovation, and industry self-regulation by large tech companies We then identify three types of structures (organizational, regulatory, technological/algorithmic), which synergistically could help enact needed regulatory oversight while limiting the friction and economic costs of regulation. This analysis provides a starting point for further discussions and negotiations among stakeholders and regulators to do so.

Keywords: Patient-generated health data, PGHD, governance, privacy, Big Tech

Introduction

Mobile health apps and wearable monitoring devices are growing in popularity, and patient-generated health data (PGHD) – health data created and captured by or from patients or other non-clinical actors – are proliferating outside of clinical settings. PGHD include, but are not limited to, biometric data (e.g., from remote monitoring), self-reported health measures, health-related activity data (e.g., exercise, diet, sleep), symptoms, and health and treatment history. PGHD are distinguished from clinical data in that the patients are is the primary capturers of their own data, and patients usually can decide how, and with whom, to share these data (Office of the National Coordinator for Health Information Technology, 2014). In clinical settings, PGHD may supplement data collected by doctors. Existing examples of PGHD applications include sleep trackers, fitness trackers (e.g., Fitbit and Apple Watch), fertility apps, continuous glucose monitors, smart thermometers, EKG monitors (e.g., AliveCOR), and consumer DNA tests to examine health predictors (e.g., 23andMe), with many additional applications existing or in development. Individuals are opting to use these technologies for a variety of reasons, including self-monitoring for health maintenance or improvement, the intention to link PGHD to physicians for continuity of care, or out of altruistic concern for public welfare (e.g., sharing health data to track COVID-19 or for medical research).

Most PGHD technologies do not undergo FDA scrutiny for efficacy or accuracy: "the availability and development of the technologies has, in many instances, outpaced the publication of trials designed to evaluate health outcomes, usability, interoperability, and benefits and harms of these technologies" (Agency for Healthcare Research and Quality, 2019, p.1). In addition to concerns about safety, accuracy, and efficacy, the governance of personal, even

intimate, health data collected by, stored, and shared by these devices is an important area of concern (e.g., Montgomery, Chester & Kopp, 2018).

Where and how PGHD are collected, aggregated, and maintained vary. The technology firms that provide PDHD mobile applications or monitoring devices often initially aggregate these data on their own IT infrastructure "in the cloud," patients may store data locally on a personal device, or both may occur. Patients may sometimes share PGHD with clinicians, directly or via the technology vendor's or clinician's health data portal. Clinicians may, or may not, integrate these data into the patient's electronic health record (Genes et al., 2018; Abdolkhani, Gray, Borda & DeSouza, 2019). The extent to with PDHG (anonymized or not) are shared with other third parties, for instance, with employers who sponsor workplace wellness programs, insurers who embed PGHD monitoring apps in their insurance policies (Aetna, 2020; Bari & O'Neill, 2019), or retailers or marketing firms, is unknown. Thus, in addition to integration (or planned integration) into clinical health settings, these data are undoubtedly being included in growing stockpiles of personal health data that may be mined for a multitude of purposes by third parties for insight via big data analytics and artificial intelligence/deep learning

Dominant narratives about big data health analytics position these highly personal data as valuable resources that are poised to transform healthcare, stimulate innovation in medical research, and provide feedback, autonomy, and control related to personal health to individuals, i.e., the Quantified Self movement, with personal health behavior being shaped by health analytics to "nudge" individuals towards certain actions (Swan, 2013). Governing the vast array of personal health data to facilitate patient care and health research while preserving individual privacy and autonomy will be challenging, even more so with PGHD, which are the least regulated domain of digitalized personal health data (U.S. Department of Health and Human

Services, 2018). Concerns about how these data could lead to unjust discrimination through differential pricing, increased health premiums, or denial of health insurance – or that the benefits will enrich only the lives of a privileged few – are emerging (Winter & Davidson, 2019).

In this paper, we explore three existing data protection regimes relevant to PGHD in the United States that are currently in tension with one another: federal and state health-sector laws, data use and reuse for research and innovation, and industry self-regulation by large tech companies. We examine and question the extent to which an enhanced federal health privacy law or stricter omnibus data protection laws more akin to GDPR (Davis, 2020) will be enough to address this new environment. We conclude by outlining three related categories of governance structures that must be harmonized as stakeholders' values and interests with regards to personally generated health data diverge: organizational, regulatory, and technical/algorithmic.

Health data governance: Three data protection regimes relevant to PGHD

Most discussions of data governance related to personal health information center on the important concerns of privacy and security. In the United States, a deeper discourse on health data governance that considers what types of value are potentially afforded by personal health data and whose values and interests shape governance structures and goals toward realizing value has not yet emerged. This is due in large part to the assumption that data collected by an organization are de facto the assets (IP) of that firm, to govern and to utilize as they deem appropriate within applicable regulatory regimes (Winter & Davidson, 2017). Below, we review three existing data protection regimes relevant to PGHD and governance issues within each.

U.S. Federal and State health-sector laws (regime 1)

When considering the U.S. approach to privacy and data protection, it is important to acknowledge that it is an ambiguous concept with no consensus definition, and it has evolved over time (Solove, 2010; Acquisti, Brandimarte, and Loewenstein, 2015; Igo, 2018). The U.S. has not had a major federal privacy reform in several decades, relying instead on a loose collection of sector-specific laws. Additionally, states have their own laws, and where these offer more protection to the individual, they often take precedence over federal law.

In the healthcare sector, the Health Insurance Portability and Accountability Act of 1996 (HIPAA) mandated the protection and confidential handling of certain health information (protected health data) and addressed digital data developments evident at that time. HIPAA required the U.S. Department of Health and Human Services (HHS) to develop regulations and accountability mechanisms. HHS published two related rules. The *Standards for Privacy of Individually Identifiable Health Information* ("Privacy Rule") set national standards for data protection of some health information. The *Security Standards for the Protection of Electronic Protected Health Information* ("Security Rule") set national security standards for certain data in digital form. Within HHS, the Office for Civil Rights (OCR) is tasked to enforce both rules. Additional provisions to the HIPAA privacy and security regulations were introduced as part of the Health Information Technology for Economic and Clinical Health Act (HITECH) in 2009, and in 2013 the HIPAA Omnibus Rule updated these privacy and security regulations. (U.S. Department of Health and Human Services, nd). Other narrow slices of health data are also protected under the Fair Credit Reporting Act (limiting data use and sharing of medical information in consumer financial reports), the Genetic Information Nondiscrimination Act of 2008 (prohibiting discrimination in health insurance and employment based on genetic information), Family Educational Rights and Privacy Act (limited sharing of medical

information found in school records), and a few other laws related to personal health information held by the government.

HIPAA applies only to specified covered entities (e.g., physicians, hospitals, insurers) and their designated business associates. As a result, the health data regulated via HIPAA originate within or are related to clinical healthcare settings. For instance, physicians' notes in an electronic health record (EHR), prescriptions or laboratory orders exchanged with pharmacies or labs, insurance claims for reimbursement, health information exchange between clinicians, data in EHR-tethered patient portals, and so on, fall under the regulatory regime of for health data and specifically, HIPAA. HIPAA applies to PGHD collected by patients only in certain limited circumstances, for instance if a physician prescribes to a patient a specific health monitoring device provided by a business associate, which then manages data on behalf of the clinician.

In the decades since the law was created, consumers' widespread use of the Internet for health information, proliferation of Internet of Things devices (including mobile devices and health monitors), digitalization of business and government transactions and records (such as Electronic Health Records and insurance claims filing), and advanced user profiling and data analytics software have led to an environment where a great deal of health-related data not covered by HIPAA or other laws are collected and shared. Since PGHD are collected by patients themselves using mobile apps provided by technology firms and are typically collected outside clinical encounters, these data fall outside of HIPAA (U.S. Department of Health and Human Services, 2016, 2018). Instead, "PGHD are maintained primarily on the information technology (IT) infrastructure of vendors who provide the mobile apps and devices, and data are governed under the IT firm's own privacy policies and within the firm's intellectual property rights in data" (Davidson, Winter & Fan, 2019, p.1). Patients can view and may be able to download their

data to devices they control, and they may be able share data with clinicians, thus replicating data in multiple locations. If PGHD are integrated into clinical EHR systems managed by covered entities, the instance of integrated data would then fall under HIPAA regulations.

Data reuse for research and innovation (regime 2)

The promises of transformational improvements in health care treatments, as well as in the cost, quality, and access to healthcare services, depend on access to health data, particularly individual-level health data, and on technology innovations associated with data analytics and artificial intelligence. Researchers in many domains (health, life sciences, technology development, etc.) are anxious to obtain access to protected health data, and increasingly, to PGHD. Their access to and sharing of health data are subject to privacy-preserving regulations, as well as to hurdles such as competitive IP interests of data-holding organizations (Rosenbaum, 2010).

In the U.S., health data used for research may be governed by general research regulations for ethical treatment of research subjects (U.S. Department of Health and Human Services, 2016). The "Common Rule" 45 CFR 46 specifies that research conducted by federal agencies or institutions receiving federal funding be reviewed and approved by Institutional Review Boards (IRBs), which assess the scope, research value, and prospective regulatory compliance of human subjects research. Due to the broad reach of federal funding, the vast majority of university-affiliated and health system-affiliated researchers fall within this scope. When such research involves health data, additional regulations may apply. In particular, HIPAA privacy rules apply to all protected health data (discussed above), along with specific privacy protections such as 42 CFR Part 2 for data on substance abuse. State government agencies and

other health data holders may also impose additional regulations on the access to and uses of health data under their governance.

The route most researchers take to obtain individual-level protected health data depend on whether the data are anonymized or are identified (or identifiable). HIPAA specifies 18 data fields (name, DOB, medical record number, phone number, dates medical services were provided, and so on) that must be removed or masked to be fully "deidentified". Once data in a dataset are deidentified, data can be provided as a public use file (PUF) not subject to HIPAA, per se. Requests for fully or partially identifiable data are subjected to some form of IRB or other approval authority and typically are also subject to data use agreements that specify how data can be used and when data must be disposed of. Policies and practices for granting access to protected health data from government agencies such as Medicare claims data from HHS/ CMS, are developed through administrative rules making and are published publicly.

In contrast to these formal procedures and policies for protected health information, how and when PGHD are being (or should be) used in research are much less clear. As noted earlier, PGHD are generated in a wide variety of socio-technical contexts, many which are not subject to HIPAA or related regulation. If researchers subject to the Common Rule receive identifiable PGHD directly from research subjects, then their data use will be subject to 45 CFR 46, but not to HIPAA, since these data are not covered by HIPAA. If these researchers obtain PGHD directly from the technology vendor versus from the research subjects, the application of 45 CFR 46 is less clear. Well-intentioned attempts to anonymize such data would likely fail to achieve that goal, since PGHD contain many subtle, identifying attributes not specified by HIPAA criteria. However, individually-identifiable data that are derived from pre-existing files without the researcher's interaction with the subjects of that data, can fall outside the scope of "human

subjects research". Whether use of PGHD would fall under the Common Rule in these circumstances would depend on the individuals' expectations for privacy on the technology vendor's IT platform. Expectations of privacy then depend on the vendor's stated privacy policy. Importantly, PGHD technology vendors selling products and services directly to consumers (vs. as a business associate for a covered entity) likely are not be subject to either 45 CFR 46 or HIPAA regulation. Thus, in most instances, how PGHD are used in research will depend on the self-regulation of technology vendors in this market.

Big tech self-regulation (regime 3)

Big Tech companies are racing to enter the lucrative healthcare sector, as providers of secure cloud services and AI-enabled analytics to health care organizations, developers of a variety of consumer-based health apps, and in medical research. While companies like Google, Apple, and Microsoft's earlier forays into healthcare were slow to take off or failed (Davidson, Østerlund, & Flaherty, 2015) over the past several years, these companies have reorganized and refocused their health initiatives while being joined by IBM, Amazon, Facebook, and others. For instance, Apple's healthcare efforts now center on three main areas: 1) consumer products based on wearable and smartphone apps; 2) health research in partnership with health organizations enabled via Apple's open-source framework, ResearchKit; and 3) secure health records via its Health app, which was launched in 2014 and consolidates data from an individual's phone, watch, and health-related third-party apps (Dyrda, 2019). In 2020, Google Health has grown to over 500 employees, with many allocated from other parts of Alphabet (Farr, 2020), including its health AI venture, Google DeepMind Health. Google also bought leading fitness tracker Fitbit in late 2019.

The regulatory regimes that govern Big Tech and consumer privacy differ substantively from HIPAA-related regulations that oversee protected health data. Between 1997 and 2007, there was an array of efforts to create industry- or government- supported self-regulatory guidelines for handling of personal data, but "the majority of the industry self-regulatory programs that were initiated failed in one or more substantive ways, and, many disappeared entirely" (Gellman & Dixon, 2011, p.2). Today, federal and state levels require companies to have and enforce a privacy policy if they gather and use personal data. The Federal Trade Commission (FTC) serves as regulatory authority for many of these laws and requires specific language and provisions to ensure privacy. However, aside from these stipulations, there is little control over policy content, and data subjects may not fully read or understand the agreements and the potential consequences of sharing data. These agreements often state that data may be shared with third parties. However, audit and enforcement are lax, and when a firm is absorbed by another via sale, previous privacy policies may be nullified. For instance, when popular fitness tracking app Fitbit was purchased by Google in 2019, there was concern about how Google would use the PGHD accumulated via Fitbit, since "current laws and regulations do little to hold Google and other companies to their promise" (Frazee, 2019, para. 4).

As these firms compete to gather PGHD and other data that infers health information, their efforts often bypass existing health regulation in the U.S. Most PGHD devices are not classified as medical devices by the Food and Drug Administration, and the self-reported or inferred health data collected and aggregated by Big Tech also does not fall under HIPAA. When consumers (patients) transfer protected health data from a HIPAA-regulated context (such as their clinician's patient portal) to the technology firm's PGHD infrastructure (such as Apple's Health Kit), these instances of their PHI data are no longer HIPAA-protected. These large tech

firms also have access to detailed information about an individual's search history, social networks, interests, and consumer consumption patterns that are not explicitly health-related but can reveal aspects of a person's health. For example, information about a person's location over time, their associates, or demographic and community data may be used to infer health. Essentially, "all your data is health data" through a process called digital phenotyping, which refers to "taking information from our digital behaviors — on websites, via our phones — and using it to gain insight into potential health issues" (Warzel, 2019, para. 4).

Tensions are developing between suspicions about Big Tech's intentions for PGHD and other health data and health care innovations through advanced analytics and AI technologies these firms might provide. Particularly concerning are the lack of transparency in data use agreements between HIPAA-regulated entities and Big Tech and the data use policies that Big Tech will craft and adhere to related to reuse of PHI across an array of AI development projects. For instance, in November of 2019, a whistleblower working at Google (Anonymous, 2019; Pilkington, 2019) revealed that Google Health was partnering with a large health system nonprofit, Ascension Health, to analyze the PHI of approximately 50 million people across 21 U.S. states. Called Project Nightingale, this partnership raised immediate public, health industry, and governmental concern (Copeland, 2019; Copeland & Needleman, 2019; Loveland, 2020; Pifer, 2019; Price, 2020), and the Department of Health and Human Services Office of Human Rights opened an investigation to determine whether HIPAA violations had occurred. Some members of Congress also expressed grave concerns about the partnership and called for both parties to disclose their data use agreement (United States Senate, 2020).

It appeared that both parties were technically compliant with HIPAA, as Google was designated as a business associate providing the IT infrastructure and records interface for

Ascension. However, questions about Google's plans for using this protected health data to develop AI capabilities remain. Given Google's access to vast storehouses of PGHD from Fitbit, mobile apps, and other sources that could be added to such clinical data, the potential for dramatic innovation but also devasting losses of personal privacy and autonomy are evident in the Nightingale case. Google is not alone in its aspirations, as other Big Tech firms are undertaking similar ventures.

Towards harmonizing regulatory regimes for PGHD governance

Governing PGHD to facilitate innovation (e.g., improved medical care, energy efficiency, scientific advances) while also preserving individual privacy and autonomy will be challenging. In the past, health data governance was largely handled within health industry organizations with well-defined regulatory responsibilities. With the rise of PGHD, along with the many partnerships among health industry organizations and Big Tech, health data governance responsibilities are now shared among numerous types of institutions and firms, which operate in different regulatory regimes for data governance and with different norms and practices for protecting privacy and for sharing data.

Figure 1 highlights the governance challenges PGHD pose, as data arise within or shift across different regulatory boundaries. Individuals, whether in their roles as patients or simply as consumers, bear some responsibility for understanding what data they are generating and deciding what data uses are acceptable and in what circumstances. To do so, they need clear data use and privacy policies as well as accessible methods to enact their informed consent. HIPAA-regulated data stewards and researchers need clarity on what regulations apply to their own use of PGHD and in which circumstances, as well as what responsibilities they have for data they

provide to or acquire from other organizations. Technology vendors venturing into healthcare in pursuit of great innovation and even greater profits have a duty to fully document and monitor PGHD use in compliance with applicable data regulations. Their leaders must understand and accept the differences between self-regulation and expectations for ethical health data governance, particularly in relation to health analytics and AI algorithms. Both HIPAA-regulated and technology firms must balance their goals of protecting IP and market share from innovation with a high degree of transparency in their partnerships, data use agreements, and data uses.

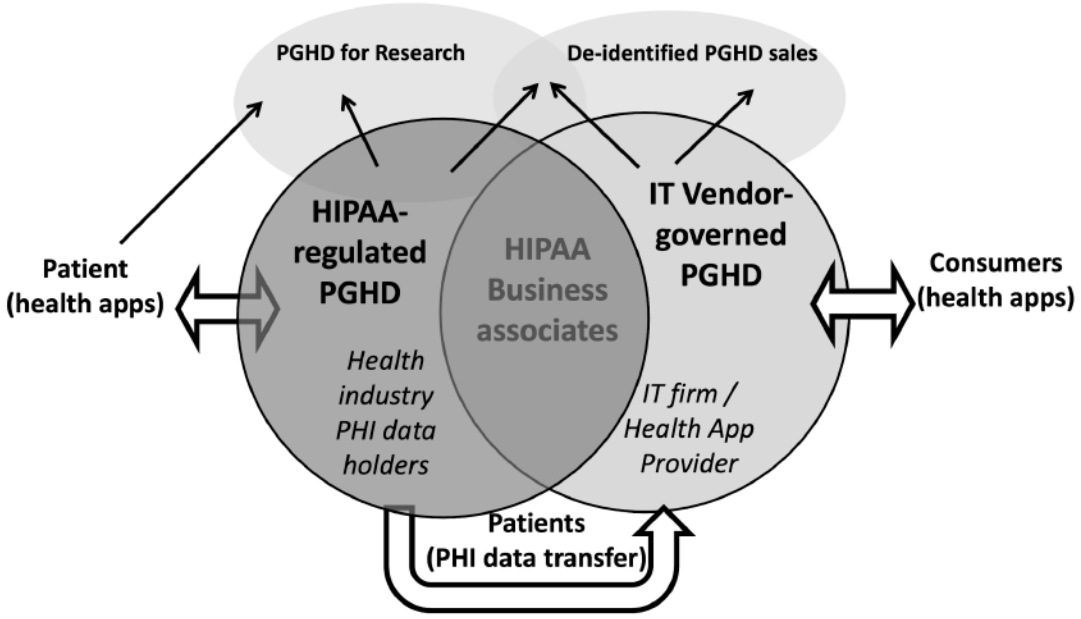

Figure 1: PGHD Governance Across Regulatory Regimes

While healthcare providers are bound to adhere to health privacy laws, Big Tech has operated under a more relaxed regulatory regime that will facilitate monetizing PGHD and creating new intellectual property with big data and AI developments. Effective governance requires surfacing conflicts between diverse stakeholders, assessing normative goals across sectors, and determining mechanisms to detect and prevent harm and enforce

compliance/accountability. Data governance strategies must balance interests of multiple stakeholders with different expectations, accountabilities, and ethical codes arising under different regulatory requirements.

We refer to these processes as *harmonization of regulatory regimes*. The lack of harmonization between regulatory regimes may render ineffective the existing policies and interventions to preserve patient privacy and control over their health data generally, and PGHD in particular. A common refrain is that we need stricter data protection laws more akin to GDPR, but will these be enough to address this new environment? To address this question, we consider three overlapping categories of data governance structures that must be harmonized as stakeholders' values and interests with regards to PGHD diverge: Organizational, Regulatory, and Technical/Algorithmic (Figure 2). We provide examples of emerging governance structures for each and consider how these structures are mutually dependent on the others.

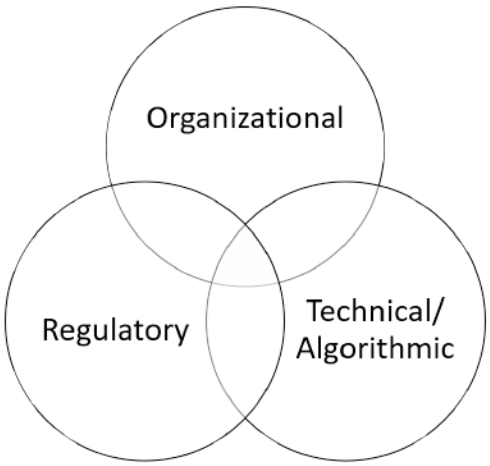

Figure 2. Three overlapping categories of data governance structures

Organizational structures

A variety of organizational data governance structures have emerged to address the challenges of digitized health information. One example is data trusts (Delacroix & Lawrence, 2019; O'hara, 2019), agreements where data subjects agree to pool their rights to achieve common goals that require the use of the pooled data. A fiduciary, or data trustee, makes agreements for use of the data on behalf of the data contributors. "This 'bottom-up' data trust model is resolutely complementary to top-down, regulatory constraints (including those of the GDPR)" (Delacroix & Lawrence, 2019, p. 243). Data trusts are thus a potential organizational solution to some multi-stakeholder problems by holding user data and ensuring acceptable use, outside any one organizational boundary. Data trusts could foreseeably enable individuals to benefit financially from their PGHD (and other health data) contributions, rather than all value from data sales accruing to the technology vendor. However, these arrangements do not negate the need to negotiate, document and effectively communicate data use policies both to those who contribute their data and to potential users of the data. For instance, data contributors would need to understand how their data would be used and if they could later revoke their consent, even after data have been provided to researchers or other users. Moreover, it will be difficult for a data trust to ensure that data use agreements are complied with or to take preventative actions against misuse, if data are replicated for use outside of the IT infrastructure controlled by the trust.

A second example of an emerging organizational governance structure is the "AI ethics board". Many companies have established these as internal groups that provide governance, recommendations, and oversight regarding ethical AI development in the organization. In some cases, organizations have also tapped external review committees to review their work and

provide assurance to the public and regulators that the company is behaving ethically. Doing so effectively would be a step towards greater transparency on health data and PGHD use. However, deciding who should appoint such boards, who the boards are accountable to, and what latitude boards have will be key to their effectiveness. Whether such boards are established and how they operate thus might fall under a regulatory regime similar to the Common Rule and IRBs, rather than remaining in the self-regulatory regime that Big Tech currently enjoys.

For instance, after widespread public outcry about its data sharing relationship with the UK National Health System in 2016, Alphabet's (Google's) DeepMind Health created two entities: 1) an external Independent Review Panel tasked to review DeepMind Health's activities and issue an annual report to the public; and 2) the DeepMind Ethics & Society Fellows to research issues such as privacy and fairness (Winter & Davidson, 2018). However, rather than engendering greater trust, this latter group was perceived as a "gigantic conflict of interest" due to "a commercial AI giant researching the ethics of its own technology's societal impacts" (Lomas 2017, para. 5). During the merger of DeepMind Health and Google in 2019, Google also created the Advanced Technology External Advisory Council (ATEAC) to guide ethical development of new technologies. ATEAC was shut down within a week due to controversial choices of members and principles that were hard to enforce, "not least because the enforcement mechanisms for violating the principles aren't well-defined, and, in the end, the entire enterprise remains a self-regulatory endeavor" (Johnson & Lichfield, 2019, para. 7). Concerns about such efforts by Big Tech to self-manage these issues have recently come to public attention with the controversial exit of Google AI ethics researcher Dr. Timnit Gebru, who had co-authored a paper under review that emphasized several risks of large language models, which are central to Google's research and innovation (Hao, 2020).

Regulatory structures

Regulatory data governance structures are intended to protect health data privacy and security while not unduly posing stumbling blocks for innovation. In the U.S., proposed regulatory governance structures include an update to HIPAA focusing on expansion of covered entities, informed consent, and research use. One action would be to expand HIPAA to cover at least some domains of PGHD (Bari & O'Neill, 2019). A key limit of HIPAA today is that it only applies to "business associates" of the originating healthcare actor. For instance, in Google Health's partnership with Ascension, the former is classified as a business associate to the later for HIPAA purposes, but when Google Fitbit sells a fitness tracking device and an individual uses it, data generated by the device are not covered entities by HIPAA.

Expansive uses of PGHD and other health data for health research are critical to innovation, but such uses can challenge informed consent policies, because it is not possible to forecast all possible uses of health data or the risks posed at the time data is collected (Sharon, 2016). Moreover, keeping track of where and how health data are used from the point of origin through the many instances of use will require greatly enhanced documentation and tracking capabilities to enable renewed consent and audits for compliance. New models of broad and portable consent for large-scale research projects using health data are being explored to address "the unavoidable question of who stands to benefit and in which way from research results" (Sharon, 2016, p. 568).

The U.S. might also move from a dependence on sectoral data regulation towards a new omnibus data protection law. The EU's General Data Protection Regulation (GDPR) has stimulated a global discussion about data privacy and protection and many jurisdictions are

moving towards GDPR-compatible regimes. The EU's GDPR, which was enforced in May 2018, and the passage of California's Consumer Privacy Act (CCPA) in June of 2018 (enforced in 2020), have renewed discussion about an omnibus U.S. federal privacy law. Big Tech firms, under increasing federal and international scrutiny, have also called for legal reform: "In November 2018, in response to a call for comments on a federal privacy law by the NTIA, numerous companies responded by stating that they were now in favor of a federal privacy law" (Solove, 2019, para. 3). However, at this time, key uncertainties will likely shape the outcome of U.S. federal privacy law for the near future: (i) The resolution of tensions between state and federal privacy laws given the current Congressional deadlock and new administration; (ii) the possibility that stricter state laws may be preempted by a less stringent federal general privacy law; (iii) the fallout of the Privacy Shield decisions related to U.S.-European personal data sharing and other GDPR enforcement, and (iv) pending antitrust actions towards large tech companies (e.g., Federal Trade Commission v. Facebook, Inc., 2020).

Another approach is the creation of regulatory sandboxes, that is, supervised environments managed by a data protection authority to pilot innovative products or business models in a live market with real consumers. Each participant has clear goals (e.g., improved health outcomes) and tests on a small scale with limited time and consumers (Centre for Information Policy Leadership, 2019). With regards to data protection, a regulatory sandbox "can simultaneously address two inevitable uncertainties—the uncertainties of innovation ('what is this going to deliver?') and the uncertainties of principles-based regulation ('will this processing be fair?')" (Centre for Information Policy Leadership, 2019, p.5). This allows companies to test out new ideas without concern for regulatory challenge and enforcement, and individuals benefit from more scrutinized and customized privacy safeguards. Such approaches

are promising to avoid overregulation as well. Regulations are not frictionless; they can stifle innovations in health care research, services, and products. Thus, evidence-based regulatory evaluations such as through a sandbox approach can help determine whether a new regulation is meeting its goals at acceptable socio-economic costs and risks.

Technological/algorithmic structures

Technological/algorithmic governance structures are also emerging to address the challenges of digitized health information and growth in PGHD. Data use and movement often lacks transparency to data subjects or regulators, and thus noncompliance and associated harms are hard to detect. Metadata management software tools exist today that can identify data types (such as PHI) in data stockpiles, reverse engineer data structures and trace data transformations across systems, and facilitate analysis of data lineage. If such metadata incorporates regulatory compliance rules, then systems developers can avoid building non-compliant systems, and auditors can more readily assess and monitor compliance. However, acquiring, implementing, and maintaining metadata management software will require significant resources, and not all PGHD firms will volunteer to apply these practices without regulatory compliance incentives.

Concern about transparency in data use agreements has led to a movement towards fairness, accountability, and transparency (FAT) in algorithms. One example of this is establishing audits via blockchain, proposed as a means to ensure repudiation of transactions does not happen and that transactions cannot be altered at a later date. For instance, as a result of regulatory censure and a loss of public trust over the DeepMind Health-NHS partnership, DeepMind announced it was using an automated audit of health data accessed using a new blockchain-like technology called the "Verifiable Data Audit" (DeepMind 2017f; Hern 2017).

This new governance structure purportedly provides a real-time audit and verification of data access and use (Winter & Davidson, 2018).

Blockchain technology has also been proposed as a means for individuals to manage their own medical records and data. "Because blockchain maximizes security and accessibility, the technology can be used in many different areas of the healthcare system, such as for storing and sharing medical records and insurance information both in healthcare venues and in mobile applications and remote monitoring systems, and for clinical trials" (Chen et al., 2019). In 2016, the government of Estonia established a national cloud-based, blockchain-secured electronic health records system to secure personal health data while making it available to the individual and his or her healthcare providers (Park, 2019).

Such technological developments that might enhance the efficacy of regulatory control are sorely needed. Google's continued forays into clinical health data acquisitions demonstrate that the allure of profit from development of AI intellectual property may outweigh concerns about regulatory consequences. Even in a highly regulated environment such as the UK, where Google DeepMind's partnership with the UK National Health Services led to public and regulatory censure (Powles & Hodson, 2017), Google (via its DeepMind Health operation, which has since been absorbed into Google Health) continued to acquire patient data (Winter & Davidson, 2018). In 2019, the University of Chicago Medical Center and Google were also sued over their partnership to share personal health data (Wakabayashi, 2019).

Further, in a global race for AI dominance across sectors, countries with fewer restrictions on data use, such as China and Russia, have a notable advantage due to access to large data sets. In healthcare, the rapid deployment of AI to address the COVID-19 pandemic,

including successes in vaccine development, drug discovery, and disease diagnosis and monitoring (National Institutes of Health, 2020; Harmon et al., 2020; Arshadi et al., 2020) has been lauded and may accelerate or legitimize the desire for fewer restrictions on health data use. COVID-19 tracking apps that generate PGHD with a range of data use permissions have been deployed.

Finley (2020) sees potential for collaborations between Big Tech and smaller companies with enterprise-focused privacy platforms. As deep learning algorithms become more sophisticated, disparate sources of data may be linked to enhance predictive analytics (Bates et al., 2014; Siegel, 2016). Big tech firms rely on these vast amounts of data from disparate sources to develop large language models, a core part of their business, and PGHD are a lucrative target. PGHD "can be combined with personal information from other sources– including healthcare providers and drug companies–raising such potential harms as discriminatory profiling, manipulative marketing, and data breaches" (Montgomery, Chester & Kopp, 2018, p. 42). Assessment of whether firms are compliant with health laws is becoming increasingly difficult due to AI/deep learning's opacity (Winter & Davidson, 2019).

Thus, another example of a promising technological/algorithmic governance structure is "explainable AI". "To bring deep models built from EHR [electronic health records] data into real use, users often need to understand the mechanisms by which models operate. Such a level of model transparency is still challenging to achieve" (Xiao et al., 2018, p. 1425). Explainable AI employs mathematics to simplify the so-called black box so that humans can understand the path an AI took to reach an outcome (Abdul et al., 2017; National Institutes of Standards and Technology, 2020), thereby illuminating how personal data including PGHD were used (or misused) .

Conclusion

The growing volume of personally generated health data reflects innovations in health care that can help individuals better manage their health, healthcare providers to advise and treat their patients, third-party payers to incentivize subscribers, and providers to improve health outcomes and health service efficiency. Technology vendors, and particularly the Big Tech firms that have brought about transformative innovations in other socioeconomic sectors, are well positioned to similarly transform healthcare and to generate substantial profits as they do so. Despite some shared values related to improving health, these various stakeholders have different goals and priorities for how PGHD should be governed and used, and how and by whom value arising from PGHD should be captured. The norms and practices that varied stakeholders adopt, and actually practice (verses espouse), also differ substantively. Effective regulations are critical public policy tools to help balance stakeholders' interests.

In this paper we have argued that the diverse regulatory regimes that stakeholders operate within contribute to confusion about what regulations apply to PGHD in different contexts and what compliance may entail. There are lacunae that will enable misuse of PGHD if not addressed. Existing health data regulations are not fully applicable to PGHD, so that further elaboration of existing regulations such as HIPAA is needed, and new forms of regulation may be required. On the other hand, there may be instances where regulatory regimes overlap and are redundant, and thus may stifle valuable innovations. We have advocated that these regulatory regimes be examined and harmonized with respect to governance of PGHD. We identified three types of structures (organizational, regulatory, technological/algorithmic) and provided examples within each type, which synergistically could help enact needed regulatory oversight while

limiting the friction and economic costs of regulation. This analysis provides a starting point for further discussions and negotiations among stakeholders and regulators to do so.

References

Abdolkhani, R., Gray, K., Borda, A., & DeSouza, R. (2019). Patient-generated health data management and quality challenges in remote patient monitoring. *JAMIA Open*, *2*(4), 471-478.

Abdul, A., Vermeulen, J., Wang, D., Lim, B.Y. and Kankanhalli, M. (2017). Trends and trajectories for explainable, accountable, and intelligible systems: an HCI research agenda. CHI 2018.

Acquisti, A., Brandimarte, L., and G. Loewenstein. (2015). Privacy and human behavior in the age of information. *Science, 347*(6221), 509-514.

Aetna. (2020). Employee health and wellness programs. Retrieved from: https://www.aetna.com/insurance-producer/health-wellness/wellness-programs.html

Agency for Healthcare Research and Quality. (2019, Dec. 3) Automated-entry patient generated health data for chronic conditions: The evidence on health outcomes. Evidence-based Practice Center Technical Brief Protocol. Revised March 17, 2020.

Anonymous. (2019, November 14). I'm the Google whistleblower. The medical data of millions of Americans is at risk. *The Guardian*. Retrieved from: https://www.theguardian.com/commentisfree/2019/nov/14/im-the-google-whistleblower-the-medical-data-of-millions-of-americans-is-at-risk

Arshadi A K, Webb J, Salem M, et al. (2020). Artificial Intelligence for COVID-19 Drug Discovery and Vaccine Development. *Frontiers in Artificial Intelligence*, 3, 65.

Bari, L., & O'Neill, D. P. (2019). Rethinking patient data privacy in the era of digital health. Health Affairs Blog. Retrieved from: https://www.healthaffairs.org/do/10.1377/hblog20191210.216658/full/

Bates, D.W., Saria, S., Ohno-Machado, L., Shah, A. and Escobar, G. (2014). Big data in health care: Using analytics to identify and manage high-risk and high-cost patients. *Health Affairs, 33*(7), 1123-1131.

Centre for Information Policy Leadership. (2019, Mar. 8). Regulatory sandboxes in data protection: Constructive engagement and innovative regulation in practice. Retrieved from: https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/cipl_white_

paper_on_regulatory_sandboxes_in_data_protection_-_constructive_engagement_and_innovative_regulation_in_practice__8_march_2019_.pdf

Chen, H. S., Jarrell, J. T., Carpenter, K. A., Cohen, D. S., & Huang, X. (2019). Blockchain in healthcare: A patient-centered model. *Biomedical journal of scientific & technical research*, *20*(3), 15017.

Copeland, R. (2019, Nov. 11). Google's 'Project Nightingale' gathers personal health data on millions of Americans. Retrieved from: https://www.wsj.com/articles/google-s-secret-project-nightingale-gathers-personal-health-data-on-millions-of-americans-11573496790

Copeland, R., & Needleman, S.E. (2019, Nov. 12). Google's 'Project Nightingale' triggers federal inquiry. The Wall Street Journal. Retrieved from: https://www.wsj.com/articles/behind-googles-project-nightingale-a-health-data-gold-mine-of-50-million-patients-11573571867

Davidson, E. J., Østerlund, C. S., & Flaherty, M. G. (2015). Drift and shift in the organizing vision career for personal health records: An investigation of innovation discourse dynamics. *Information and Organization*, *25*(4), 191-221.

Davidson, E., Winter, J. S., & Fan, V. (2019). Challenges and opportunities with governance of personally generated health data (PGHD). Conference on Health IT and Analytics (CHITA). Retrieved from: http://hdl.handle.net/10125/63535

Davis, J. (2019). Google Ascension Partnership Fuels Overdue HIPAA Privacy Debate. Retrieved from: https://healthitsecurity.com/news/google-ascension-partnership-fuels-overdue-hipaa-privacy-debate

DeepMind. (2017). Trust, confidence and verifiable data audit. Retrieved from: https://deepmind.com/blog/trust-confidence-verifi-able-data-audit/

Delacroix, S., & Lawrence, N. D. (2019). Bottom-up data Trusts: disturbing the 'one size fits all' approach to data governance. *International Data Privacy Law*, *9*(4), 236-252.

Dyrda, L. (2019, Nov. 8). Tim Cook: 4 key thoughts on Apple's healthcare offerings today and in the future. Becker's Health IT. Retrieved from: https://www.beckershospitalreview.com/healthcare-information-technology/tim-cook-4-key-thoughts-on-apple-s-healthcare-offerings-today-and-in-the-future.html

Farr, C. (2020, Feb. 11). Google Health, the company's newest product area, has ballooned to more than 500 employees. CNBC. Retrieved from: https://www.cnbc.com/2020/02/11/google-health-has-more-than-500-employees.html

Federal Trade Commission v. Facebook, Inc., (2020, December 9). Retrieved from: https://www.ftc.gov/system/files/documents/cases/1910134fbcomplaint.pdf

Finley, D. (2020, Nov. 17). Data permissions for coronavirus-related iOS apps vary. Retrieved from: https://www.businessinsider.com/data-permissions-for-coronavirus-related-ios-apps-differ-widely-2020-11

Frazee, G. (2019, Nov. 1). Google bought Fitbit. What does that mean for your data privacy? PBS News Hour. Retrieved from: https://www.pbs.org/newshour/economy/making-sense/google-bought-fitbit-what-does-that-mean-for-your-data-privacy

Gellman, R., & Dixon, P. (2011). Many failures: A brief history of privacy self-regulation in the United States. World Privacy Forum. Retrieved from: http://www.worldprivacyforum.org/www/wprivacyforum/pdf/WPFselfregulationhistory. pdf

Genes, N., Violante, S., Cetrangol, C., Rogers, L., Schadt, E. E., & Chan, Y. F. Y. (2018). From smartphone to EHR: a case report on integrating patient-generated health data. *NPJ digital medicine*, *1*(1), 1-6.

Hao, K. (2020, Dec. 4). We read the paper that forced Timnit Gebru out of Google. Here's what it says. *MIT Technology Review*. Retrieved from: https://www.technologyreview.com/2020/12/04/1013294/google-ai-ethics-research-paper-forced-out-timnit-gebru/

Harmon S A, Sanford T H, Xu S, et al. Artificial intelligence for the detection of COVID-19 pneumonia on chest CT using multinational datasets. *Nature Communications*, 2020;*11*(1):1-7.

Hern, A. (2017). Google's DeepMind plans bitcoin-style health record tracking for hospitals. The Guardian. Retrieved from: https://www.theguardian.com/technology/2017/mar/09/google-deepmind-health-records-tracking-blockchain-nhs-hospitals

Igo, S. E. (2018). *The known citizen: A history of privacy in modern America*. Cambridge, Ma.: Harvard University Press.

Johnson, B. & Lichfield, G. (2019, April 6). Hey Google, sorry you lost your ethics council, so we made one for you. *MIT Technology Review*. Retrieved from: https://www.technologyreview.com/2019/04/06/65905/google-cancels-ateac-ai-ethics-council-what-next/

Loveland, L. (2020, Mar. 4). US senators question Ascension on its Google collaboration Project Nightingale. Retrieved from: https://www.mobihealthnews.com/news/us-senators-question-ascension-its-google-collaboration-project-nightingale

Montgomery, K., Chester, J., & Kopp, K. (2018). Health wearables: Ensuring fairness, preventing discrimination, and promoting equity in an emerging Internet-of-Things environment. *Journal of Information Policy, 8,* 34–77.

National Institutes of Health. (2020, Aug. 5). United States. NIH harnesses AI for COVID-19 diagnosis, treatment, and monitoring. Retrieved from: https://www.nih.gov/news-events/news-releases/nih-harnesses-ai-covid-19-diagnosis-treatment-monitoring

National Institutes of Standards and Technology. (2020, Aug.). Four principles of Explainable Artificial Intelligence. Draft NISTIR 8312. Retrieved from: https://www.huntonprivacyblog.com/wp-content/uploads/sites/28/2020/08/NIST-Explainable-AI-Draft-NISTIR8312-1.pdf

Office of the National Coordinator for Health Information Technology. (2014). Patient-generated health data. Retrieved from:

https://www.healthit.gov/sites/default/files/patient_generated_data_factsheet.pdf

O'hara, K. (2019). *Data trusts: Ethics, architecture and governance for trustworthy data stewardship* (Web Science Institute White Papers, Southampton. University of Southampton 1-27.

Park, A. (2019, June 25). What the US can learn from Estonia's cloud-based, blockchain-secured EHR system. Becker's Health IT. Retrieved from: https://www.beckershospitalreview.com/ehrs/what-the-us-can-learn-from-estonia-s-cloud-based-blockchain-secured-ehr-system.html

Pifer, R. (2019, November 18). IT execs call for HIPAA overhaul in 'Project Nightingale' wake. https://www.healthcaredive.com/news/it-execs-call-for-hipaa-overhaul-in-project-nightingale-wake/567520/

Pilkington E. (2019, Nov. 12). Google's secret cache of medical data includes names and full details of millions – whistleblower. Retrieved from: https://www.theguardian.com/technology/2019/nov/12/google-medical-data-project-nightingale-secret-transfer-us-health-information

Powles, J., and H. Hodson. 2017. Google DeepMind and healthcare in an age of algorithms. *Health Technology* 7(4):351-367. doi.org/10.1007/s12553-017-0179-1

Price, L. (2020, Jan. 16). Project Nightingale: Google's four pillars for their secret patient data partnership. Retrieved from: https://www.healthcare.digital/single-post/2019/11/13/Project-Nightingale-Googles-four-pillars-for-their-secret-patient-data-partnership

Rosenbaum, S. (2010), Data governance and stewardship: designing data stewardship entities and advancing data access, Health Services Research, Vol. 45 No. 5p2, pp. 1442-1455.

Siegel, E. (2016). *Predictive analytics: The power to predict who will click, buy, lie, or die*. Hoboken, NJ: John Wiley & Sons.

Solove, D. J. (2010). *Understanding privacy*. Cambridge, MA: Harvard University Press.

Solove, D. J. (2019, April 22). Will the United States finally enact a federal comprehensive privacy law? TechPrivacy. Retrieved from: https://teachprivacy.com/will-us-finally- enact-federal-comprehensive-privacy-law/

Swan, M. (2013). The quantified self: Fundamental disruption in big data science and biological discovery. *Big data*, *1*(2), 85-99.

U.S. Department of Health and Human Services. (nd). HIPAA for professionals. Retrieved from: https://www.hhs.gov/hipaa/for-professionals/index.html

U.S. Department of Health and Human Services. (2016). Federal Policy for the Protection of Human Subjects ('Common Rule'), Retrieved from: https://www.hhs.gov/ohrp/regulations-and-policy/regulations/common-rule/index.html

U.S. Department of Health and Human Services. (2018). Conceptualizing a data infrastructure for the capture, use, and sharing of patient-generated health data in care delivery and research through 2024. Retrieved from: https://www.healthit.gov/sites/default/files/onc_pghd_final_white_paper.pdf

United States Senate. (2020, March 2). Letter to Ascension [communication]. Retrieved from: https://www.warren.senate.gov/imo/media/doc/2020.03.02%20Letter%20to%20Ascension%20re%20Project%20Nightingale%20Partnership.pdf

Wakabayashi, D. (2019, June 26). Google and the University of Chicago Are Sued Over Data Sharing. Retrieved from: https://www.nytimes.com/2019/06/26/technology/google-university-chicago-data-sharing-lawsuit.html

Warzel, C. (2019, Aug. 13). All your data is health data: And Big Tech has it all. The New York Times. https://www.nytimes.com/2019/08/13/opinion/health-data.html

Winter, J.S. (2013). Surveillance in ubiquitous network societies: normative conflicts related to the consumer in-store supermarket experience in the context of the Internet of Things. *Ethics and Information Technology*. DOI: 10.1007/s10676-013-9332-3.

Winter, J. S., & Davidson, E. (2017). Investigating values in personal health data governance models. 23rd Americas Conference on Information Systems (AMCIS). August 2017, Boston, MA.

Winter, J. S., & Davidson, E. (2018). Big data governance of personal health information and challenges to contextual integrity. *The Information Society, 35* (1), 36-51. Doi:10.1080/01972243.2018.1542648

Winter, J. S., & Davidson, E. (2019). Governance of artificial intelligence and personal health information. *Digital Policy, Regulation and Governance* (DPRG). Special issue on "Artificial Intelligence: Beyond the hype?" Doi:10.1108/DPRG-08-2018-0048

Xiao, C., Choi, E., & Sun, J. (2018). Opportunities and challenges in developing deep learning models using electronic health records data: a systematic review. *Journal of the American Medical Informatics Association, 25*(10), 1419-1428.

This material is based upon work supported by the National Science Foundation under Grant Number 1827592. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.