



# **Health Information Privacy Beyond HIPAA: A 2018 Environmental Scan of Major Trends and Challenges**

December 13, 2017

A Report for the National Committee on Vital and Health Statistics  
(NCVHS) and its Privacy, Security, and Confidentiality Subcommittee

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES

This report was prepared under contract by Robert Gellman, Privacy and Information Policy Consultant. Website: [www.bobgellman.com](http://www.bobgellman.com)

## **NCVHS Membership and Lead Staff, November 2017**

William W. Stead, M.D., **NCVHS Chair**

Bruce B. Cohen, Ph.D.

Llewellyn J. Cornelius, Ph.D.

Alexandra Goss

Nicholas L. Coussoule\*

Linda L. Kloss, M.A., RHIA,\* **Subcommittee Chair**

Richard W. Landen, M.P.H., M.B.A.

Denise E. Love

Vickie M. Mays, Ph.D., M.S.P.H.\*

Jacki Monson, J.D.\*

Robert L. Phillips, Jr., M.D., MSPH

Helga Rippen, M.D., Ph.D., M.P.H., FACPM\*

David A. Ross, Sc.D.

Debra Strickland, M.S.

Roland J. Thorpe, Jr., Ph.D.

\* Member of the Subcommittee on Privacy, Confidentiality and Security

### **Rachel Seeger, MPA, MA, Lead Staff to the Subcommittee**

HHS/Office for Civil Rights

### **Maya A. Bernstein, J.D., Subcommittee Subject Matter Expert**

HHS/Office of the Assistant Secretary for Planning & Evaluation

### **Rashida Dorsey, Ph.D., M.P.H., Executive Staff Director**

Director, Division of Data Policy

Senior Advisor on Minority Health and Health Disparities

Office of Science and Data Policy

Office of the Assistant Secretary for Planning and Evaluation, HHS

### **Rebecca Hines, M.H.S., Executive Secretary**

Health Scientist

Office of Planning, Budget and Legislation

National Center for Health Statistics

Centers for Disease Control and Prevention, HHS

**The National Committee on Vital and Health Statistics**

(NCVHS) serves as the advisory committee to the Secretary of Health and Human Services (HHS) on health data, statistics, privacy, national health information policy, and the Health Insurance Portability and Accountability Act (HIPAA) (42U.S.C.242k[k]). The Committee also serves as a forum for interaction with interested private-sector groups on important health data issues. Its membership includes experts in health statistics, electronic interchange of healthcare information, privacy, confidentiality, and security of electronic information, population-based public health, purchasing or financing healthcare services, integrated computerized health information systems, health services research, consumer interests in health information, health data standards, epidemiology, and the provision of health services. Sixteen of the 18 members are appointed by the HHS Secretary to terms of four years each. Two additional members are selected by Congress. For more information, visit the NCVHS website: <http://www.ncvhs.hhs.gov>.

## Table of Contents

<b><i>I. Introduction to Beyond HIPAA</i></b>	<b>1</b>
<b>A. Purpose and Scope of the Report</b>	<b>1</b>
<b>B. Beyond HIPAA</b>	<b>2</b>
<b>C. The Challenge of Defining Health Information</b>	<b>3</b>
1. Dilemmas and Examples	4
2. Trail of Data	9
3. Tentative Conclusions about Definitions	12
<b>D. Health Data Ownership, Control, and Consent</b>	<b>13</b>
1. The Regulated World	14
2. The Unregulated World	15
D. Fair Information Practices	17
<b><i>II. Big Data: Expanding Uses and Users</i></b>	<b>20</b>
<b>A. Overview of Big Data</b>	<b>20</b>
<b>B. Defining Big Data</b>	<b>24</b>
<b>C. Big Data and Privacy</b>	<b>27</b>
<b>D. Other Concerns about Big Data</b>	<b>30</b>
<b>E. Responses to Big Data</b>	<b>31</b>
<b><i>III. Personal devices and the Internet of Things</i></b>	<b>37</b>
<b>A. Introduction</b>	<b>37</b>
<b>B. Some Sources of Rules and Standards</b>	<b>39</b>
1. Food and Drug Administration	39
2. NIST	40
3. Federal Trade Commission	41
4. Industry and Other Standards	42
<b>C. Devices in Context</b>	<b>44</b>
1. Wellness Programs	44
2. Citizen Science	45
<b><i>IV. Laws in Other Domains</i></b>	<b>46</b>
<b>A. U.S. Privacy Model vs. the EU Privacy Model</b>	<b>46</b>
<b>B. Fair Credit Reporting Act and Its Limits</b>	<b>48</b>
<b>C. Other Sources</b>	<b>50</b>
<b><i>V. Evolving technologies for privacy and security</i></b>	<b>52</b>
<b>A. Applied technologies can get complicated quickly</b>	<b>54</b>
<b>B. Technologies can spark technical controversies</b>	<b>57</b>
<b>C. Using technology to hide data linkage</b>	<b>59</b>
<b>D. Non-Technological Protections</b>	<b>60</b>
<b><i>VI. Evolving Consumer Attitudes</i></b>	<b>61</b>

## I. Introduction to Beyond HIPAA

### A. Purpose and Scope of the Report

The purpose of this report is to provide an “environmental scan” of privacy issues for the NCVHS’s new project examining privacy and security implications of uses of health information that are outside or beyond the scope of HIPAA. The Committee commissioned this report to explore existing and emerging policy frameworks, practices, and technologies to better frame key issues and drivers of change in these areas:

1. Big data and expanding uses and users
2. Cyber-security threats and approaches
3. Personal devices and Internet of Things
4. Laws in other domains (e.g., Fair Credit Reporting restricting uses of consumer data)
5. Evolving technologies for privacy and security
6. Evolving consumer attitudes

This report does not examine issues relating to health information cyber-security threats and approaches. At the September 13, 2017, hearing, NCVHS member Jacki Monson presented the report of the Health Care Industry Cybersecurity Task Force<sup>1</sup> established by the U.S. Department of Health and Human Services (HHS) following the passage of the Cybersecurity Act of 2015.<sup>2</sup> That report reviewed and analyzed the cybersecurity challenges faced by the health care industry and set out six high-level imperatives and offered recommendations to address identified problems. The NCVHS Privacy, Confidentiality, and Security (PCS) Subcommittee decided that the report adequately covered the high level cybersecurity issues of interest to the Subcommittee in this context. This should not be read to suggest that security is a secondary concern. Indeed, as Committee Member Jacki Monson said at the November 28, 2017, virtual hearing of the Privacy, Confidentiality, and Security Subcommittee, “[i]f you don’t have security, you definitely don’t have privacy.”<sup>3</sup> The security of health data outside HIPAA is just as important as the security of HIPAA data. It just is not the subject of this report. Many of the cybersecurity principles and standards applicable to HIPAA data will have value in the world beyond HIPAA as well. Organizations and businesses processing health data outside HIPAA need to pay attention to security just as much as HIPAA covered entities.

The remaining topics covered in this report address a wide scope, and overlap to a considerable degree. For example, personal devices generate information that can become big data, use technologies that affect privacy and security, may be subject to laws outside HIPAA, and reflect consumer attitudes towards technology and privacy. We live in an interconnected world of technology and privacy, and nothing about the lack of boundaries or the overlap of policy concerns is new or unexpected.

---

<sup>1</sup> Health Care Industry Cybersecurity Task Force (HHS), Report on Improving Cybersecurity in the Health Care Industry (2017), <https://www.phe.gov/Preparedness/planning/CyberTF/Documents/report2017.pdf>. Monson was a member of the Cybersecurity Task Force.

<sup>2</sup> National Committee on Vital and Health Statistics Full Committee (Sep. 13, 2017), <https://www.ncvhs.hhs.gov/transcripts-minutes/transcript-of-the-september-13-2017-ncvhs-full-committee-meeting/>.

<sup>3</sup> Virtual hearing of the NCVHS Privacy, Confidentiality, and Security Subcommittee, (Nov. 28, 2017), <https://www.ncvhs.hhs.gov/transcripts-minutes/transcript-of-the-november-28-2017-meeting-of-the-privacy-confidentiality-and-security-subcommittee/>.

Each topic addressed here could be the subject of its own multi-volume report. The goal is to provide information to help Subcommittee members make choices about future directions for NCVHS's work and, ultimately, recommendations to the Secretary of HHS. The content reflects choices made by the author with the benefit of consultation with a modest number of experts and with guidance from the PCS Subcommittee.

## **B. Beyond HIPAA**

The focus of this report is Beyond HIPAA. That is, the report assumes the adequacy of the HIPAA Rules for health data and for covered entities subject to HIPAA, and focuses on health data outside of the HIPAA regulatory structure. The report does not address the possibility of adjustments to HIPAA statutes, rules, or guidance.<sup>4</sup>

For present purposes, the world of health data falls into two categories. Protected health information (PHI) defined by and subject to HIPAA falls in one category. The second category includes health data that does not enjoy the protections of HIPAA. **For ease of reference, the two categories are identified at times here as *regulated* (subject to HIPAA) and *unregulated* (not subject to HIPAA). Data in the unregulated category, for the most part, is not subject to any specific statutory regulation for privacy.**

Many but not all of the activities in the non-HIPAA category involve organizations that rely on health data as an element of a commercial activity, including data brokers, advertisers, websites, marketers, genetic testing companies, and others. The unregulated category includes some governmental and non-profit activities as well. The size of the unregulated world of health data is hard to estimate, but one health media expert said that in 2016, there were more than 165,000 health and wellness apps available through the Apple App Store alone.<sup>5</sup> Those apps represent a small fraction of the unregulated health data sphere.

Under HIPAA, PHI remains subject to controls in the hands of covered entities. When disclosed outside the HIPAA domain of covered entities, HIPAA data is no longer subject to HIPAA controls, although some disclosed data may occasionally fall under the scope of another privacy law. In general, however, the data disclosed by a HIPAA covered entity passes into the second category of unregulated data.

Unregulated data that passes from an unregulated actor to a HIPAA covered entity becomes PHI in the hands of the covered entity while remaining unregulated in the hands of the originator. PHI that passes out of the regulated world generally becomes unregulated data in the hands of a recipient who is not a HIPAA covered entity. Data can pass back and forth between the two worlds.

---

<sup>4</sup> This report generally ignores state laws. While some state health laws follow HIPAA boundaries, some do not. The subject is too large and complex for consideration here.

<sup>5</sup> Rice University, Press release, Be concerned about how apps collect, share health data (Oct. 19, 2017) (citing Kirsten Ostherr, Professor of English), <http://news.rice.edu/2017/10/19/rice-expert-be-concerned-about-how-apps-collect-share-health-data/>.

The focus here is on data in the unregulated world. But the borders between the two worlds may be disappearing. A recent report from a public interest group addresses the effect of the big data digital marketplace on the two worlds:

The growth of this new health economy is further eroding the boundaries between the health-care system and the digital commercial marketplace.<sup>6</sup>

Health data, whether it originates entirely in the commercial, unregulated sphere, or “leaks” into commercial databases from HIPAA regulated world, can remain essentially forever in files of data brokers and other consumer data companies. Health data may remain valuable for the entire lifetime of the data subject, and it may have uses with respect to relatives of the data subject. For example, an individual found to have a genetic condition may share the genes for that condition with relatives and children. No matter where it originated or is held, data may not be current, accurate, or complete.

The organization of the report follows the key issues identified by the Subcommittee. This introduction includes a discussion of several cross-cutting issues that should help in figuring out how to approach the key issues. These issues are: 1) The Challenge of Defining Health Information; 2) Health Data Ownership, Control, and Consent. The introduction ends with a short description of Fair Information Practices, a widely used core set of privacy principles that HHS used as a framework for the HIPAA privacy rule.

### **C. The Challenge of Defining Health Information**

If you want to discuss health data, you need to know what it is. If you want to regulate it or the institutions that have health data, you need to be able to draw clear lines. If you want to address recommendations to health data holders, you also need to be adequately descriptive. Defining health information in a broad context presents major difficulties, but not for HIPAA. HIPAA does a fine job in drawing clear lines. The real difficulties arise when you move beyond HIPAA.

HIPAA defines health information by using a series of nested definitions (health information, individually identifiable health information, and protected health information).<sup>7</sup> HIPAA ties these definitions to the covered entities regulated under the rules. The result is that HIPAA effectively covers all identifiable information related to health care treatment or to payment for the provision of health care held by a covered entity. There is no need to pick and choose among items of data to make a decision about what is and is not PHI. In practical terms, all identifiable data processed by a HIPAA covered entity is PHI subject to the HIPAA rules. Issues relating to identifiability, many already addressed recently by NCVHS, are not of immediate concern to this discussion.<sup>8</sup>

---

<sup>6</sup> Center for Digital Democracy, Health Wearable Devices in the Big Data Era: Ensuring Privacy, Security, and Consumer Protection at 21(undated, released in 2017),

[https://www.democraticmedia.org/sites/default/files/field/public/2017/aucdd\\_wearablesreport\\_final121516.pdf](https://www.democraticmedia.org/sites/default/files/field/public/2017/aucdd_wearablesreport_final121516.pdf).

<sup>7</sup> 45 C.F.R. § 160.103, <https://www.hhs.gov/sites/default/files/hipaa-simplification-201303.pdf>.

<sup>8</sup> NCVHS addressed identifiability of data in recent recommendations to the Secretary of HHS. Recommendations on De-identification of Protected Health Information under HIPAA (Feb. 23, 2017),

<https://www.ncvhs.hhs.gov/recommendations-on-de-identification-of-protected-health-information-under-hipaa/>.

HIPAA sidesteps some problems that might arise even with its broad but simple approach. It does so in part by allowing covered entities to establish hybrid entities. A hybrid entity is an organization that carries out both covered and non-covered functions.<sup>9</sup> A hybrid entity can designate the parts of its activities that fall under HIPAA and the parts that do not.<sup>10</sup> A classic example is a supermarket with a pharmacy that treats itself as a hybrid entity. The pharmacy is a HIPAA covered entity, while other activities of the supermarket are not subject to HIPAA. This separation makes it unnecessary to decide if the purchase of a can of baked beans or a bottle of aspirin creates PHI, as would happen if all of the entity's business was under the HIPAA umbrella. Everything on one side is PHI and nothing on the other side is PHI. The purchase of over-the-counter (OTC) medications from a supermarket does not create PHI. That OTC purchase begins to hint at the complexity of defining health data outside of HIPAA.

In 2007, NCVHS pointed out that a significant number of everyday providers of health care and health-related services are not covered by the HIPAA privacy and security rules."<sup>11</sup> NCVHS recommended that "HHS and the Congress should move expeditiously to establish laws and regulations that will ensure that all entities that create, compile, store, transmit, or use personally identifiable health information are covered by a federal privacy law." The issues presented by health entities within the health care system but not currently part of HIPAA (e.g., a plastic surgeon doing cosmetic procedures not covered by health insurance) are likely to be easier to address than the issues presented by the broader and more diverse class of others who process health data for different purposes, many completely external to the formal health care system. Expanding HIPAA as suggested by NCVHS to other entities within the health care system is not as difficult as identifying and defining entities outside the health care system that maintain health information.

A class of health records that can be subject to HIPAA or not subject to HIPAA is personal health records (PHRs). PHRs provided by a covered entity fall under HIPAA, while PHRs provided by a non-HIPAA entity will generally not fall under HIPAA. PHRs not subject to HIPAA can be under the general jurisdiction of the Federal Trade Commission, which has a breach notification rule for non-HIPAA PHR vendors and related parties,<sup>12</sup> but no other privacy rules for PHRs.<sup>13</sup>

## 1. Dilemmas and Examples

How can you tell what constitutes health information in the absence of a definition? This inquiry may require a determination of who has the data and for what purpose because much depends on

---

<sup>9</sup> "Covered functions means those functions of a covered entity the performance of which makes the entity a health plan, health care provider, or health care clearinghouse." 45 C.F.R. § 164.103.

<sup>10</sup> 45 C.F.R. § 164.105(a).

<sup>11</sup> NCVHS, Letter to the Secretary of HHS, Update to privacy laws and regulations required to accommodate NHIN data sharing practices (June 21, 2007), <https://www.ncvhs.hhs.gov/wp-content/uploads/2014/05/0706211t2.pdf>.

<sup>12</sup> 16 C.F.R. Part 318, <https://www.ecfr.gov/cgi-bin/retrieveECFR?gp=&SID=fd2b6cbda19e9de6a6240c59fd291989&mc=true&n=pt16.1.318&r=PART>.

<sup>13</sup> See also Maximus Federal Services, Non-HIPAA Covered Entities: Privacy and Security Policies and Practices of PHR Vendors and Related Entities Report (2012) (prepared for the Office of the Chief Privacy Officer, Office of the National Coordinator for Health Information Technology), [https://www.healthit.gov/sites/default/files/maximus\\_report\\_012816.pdf](https://www.healthit.gov/sites/default/files/maximus_report_012816.pdf).

the nature of the data and the context and the purpose. A fitness tracker may collect information about physical activity. Does the number of steps taken in a day qualify as health information? That data is PHI in the hands of a physician covered by HIPAA. There is a good argument that it is *health information* in the hands of a physical trainer at the gym. What about when a group of friends compare their fitness records among each other? What if an employer requires or encourages employees to participate in a fitness contest at the office using the tracker as evidence of activity? What if the user posts the data on a Facebook page? Does the data lose any of its character as health information if the data subject makes the data public? What if a profile of a tracker's user only shows that the user has a fitness tracker but no other information? What if the profile shows ownership and use but no details about activity? How can we tell what is *health information* once outside the traditional regulated health care context?

Here is another group of overlapping information from disparate sources. Which of these lists of individuals constitutes health information?

- Individuals formally diagnosed as obese
- Overweight individuals (by each individual's assessment)
- Big and Tall individuals (by clothing size from records of mail order and other merchants)
- Purchasers of food from a diet plan
- Purchasers of a cookbook for overweight individuals
- Individuals who visit diet websites and buy quick weight loss products (from Internet compiled profiles)

The first example, a formal diagnosis, is clearly PHI as long as the data is in the hands of a covered entity. Under HIPAA, data disclosed to a third party is no longer PHI unless the third party is another covered entity already subject to HIPAA. A disclosure to a researcher, the police, or the CIA places any PHI beyond the scope of HIPAA.<sup>14</sup> A social science researcher may obtain the data for a use unrelated to health. The police may want data to find a witness to a crime and may be uninterested in any health information about the witness. The CIA may want information for a national security purpose that does not relate to the health nature of the data. If information is not PHI in the hands of a lawful recipient, is the information nevertheless health information? And if it is, what are the rights and responsibilities of the record holder and the data subject, if any?

The above list of potentially overweight individuals returns us to the definitional question at issue here. The same information in different hands and from different sources may be health information or not health information depending on circumstances and additional information available. It is difficult to infer health information about an individual who may have bought the cookbook as a gift. Some who think they are overweight may not be (or may be anorexic).

---

<sup>14</sup> The HIPAA policy that rules do not follow records is not the only model. Under the Confidentiality of Alcohol and Drug Abuse Patient Records rules (42 C.F.R. Part 2), some records disclosed to third parties remain subject to confidentiality rules in the hands of a recipient. 42 C.F.R. § 2.32, [https://www.ecfr.gov/cgi-bin/text-id.x?SID=f561a8385178f8046a665c6c85c9b0ec&mc=true&node=se42.1.2\\_132&rgn=div8](https://www.ecfr.gov/cgi-bin/text-id.x?SID=f561a8385178f8046a665c6c85c9b0ec&mc=true&node=se42.1.2_132&rgn=div8).

The definitional problems are harder with the widespread adoption of different and better information technology and algorithms. In a now famous incident, the Target Department store used a customer profile of purchases to infer that a teenage customer was pregnant.<sup>15</sup> The customer's father protested the receipt of ads for baby products, but it turned out that the father was the last to know that his daughter was pregnant. The specific methodology Target used to make the inference is not public, but it is fair to assume that Target did not have access to health information from any HIPAA-covered source. Nevertheless, Target determined to a reasonable degree of commercial likelihood something that almost everyone is likely to agree is health information. Does the accuracy of the algorithm make a difference to whether the data is health information? If as a result of commercial or other inferences, an individual is treated as if she were pregnant, had HIV/AIDS, or had cancer, does that make the information *health information* even if the information is wrong or if the algorithm used to develop the information is highly, moderately, or not-very accurate?

Here is another example that raises the same issue from a different perspective. John Doe has an appointment at 15 Main Street at 10 am today. If a psychiatrist is the only business at that address, then the location, by itself, may be health information because it strongly suggests that Doe is a patient. If an individual, knowing of Doe's appointment, does not know anything about the office at that location, it is harder to conclude that the information is health information in the hands of that individual. If there, instead, are many medical offices at that address, with a dozen different specialties represented, then it is harder to draw specific health inferences just from the location and appointment information. If the location shared is not a street address but Latitude: N 38° 53' 9.3" Longitude: W 76° 59' 46.4," then hardly anyone would recognize the actual location without an outside reference, and there may be no practical disclosure of health information even when the address is that of the psychiatrist in solo practice.

There are more layers to the location data question. If an individual was at First and Main Streets, is that health data? What if you add the fact that the temperature that day was below zero? What if you add the fact that there was an air quality alert for that location with a reading of very unhealthy? Does it matter if the individual has emphysema? Cell phone providers collect location tracking information, with the possibility that tracking information can sometimes be considered health information, but perhaps only when someone matches an exact physical location to the cell phone owner's surroundings. For example, if the cell phone customer spends three hours, three times a week at a location occupied by a dialysis center, one might infer that the cell phone owner has kidney failure.

At the September 13, 2017, NCVHS hearing, Nicole Gardner, Vice President of IBM's Services Group, talked about the potential breadth of what information can qualify as health data. She suggested that if 50% of health is governed by social determinants not traditionally classified as health data, then nutrition, sleep, exercise, smoking, drinking, and friends may be health data.<sup>16</sup> Determining how to account for this type of data under any definition of health information is another layer of complexity. Several witnesses at the November 27, 2017 virtual hearing made

---

<sup>15</sup> Charles Duhigg, See How Companies Learn Your Secrets, New York Times (Feb. 16, 2012), <http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html>.

<sup>16</sup> National Committee on Vital and Health Statistics Full Committee (Sep. 13, 2017), <https://www.ncvhs.hhs.gov/transcripts-minutes/transcript-of-the-september-13-2017-ncvhs-full-committee-meeting/>.

similar points. Bennett Borden, partner at the DrinkerBiddle law firm, called the line between what is health-related information and what is not “very blurry.”<sup>17</sup>

Data used in one way by one party may reveal something of a health nature, but the same data held by a different party and used in a different way may not. Fitness tracker data is not PHI in the hands of a patient but it is PHI in the hands of a covered entity.<sup>18</sup> This is precisely the problem that HIPAA avoided by effectively treating all individual data as PHI if held by a covered entity.

A recent article in Slate offers a different example of the use of smartphone data (how you move, speak, type, etc.) to draw conclusions about mental health status.

That approach is fairly typical of the companies creating this new sector we might call connected mental health care. They tend to focus not on conventional diagnostic checklists and face-to-face therapy visits but on building looser, more comprehensive assess-and-intervene models based on smartphone data and largely digitized social connections. The first step in these models is to harvest (on an opt-in basis) the wealth of smartphone-generated data that can reflect one’s mental health—how you move, speak, type, or sleep; whether you’re returning calls and texts; and whether you’re getting out and about as much as usual. Such data can quickly show changes in behavior that may signal changes in mood.<sup>19</sup>

Is the information from the smartphone *health information*? Is it mental health data? None of these questions has an obvious answer.

Even health data that originated as PHI can fall outside the scope of HIPAA when held by a health data registry. Leslie Francis, Professor of Law and Philosophy, University of Utah discussed registries at the November 28, 2017 virtual hearing of the NCVHS Privacy, Confidentiality, and Security Subcommittee. Professor Francis observed that registry data often comes from clinical data subject to HIPAA, yet the data in the hands of a registry is subject to variable and incomplete privacy policies and has uneven legal protections.<sup>20</sup>

Another class of data that presents definitional and other challenges is patient-generated health data (PGHD). PGHD is “health-related data created and recorded by or from patients outside of

---

<sup>17</sup> Virtual hearing of the NCVHS Privacy, Confidentiality, and Security Subcommittee, (Nov. 28, 2017), <https://www.ncvhs.hhs.gov/transcripts-minutes/transcript-of-the-november-28-2017-meeting-of-the-privacy-confidentiality-and-security-subcommittee/>.

<sup>18</sup> For other issues related to patient generated health data, see generally Office of the National Coordinator for Health Information Technology, Issue Brief: Patient-Generated Health Data and Health IT (2013), [https://www.healthit.gov/sites/default/files/pghd\\_brief\\_final122013.pdf](https://www.healthit.gov/sites/default/files/pghd_brief_final122013.pdf).

<sup>19</sup> David Dobbs, A Sane Person’s Privacy Nightmare, Slate (Sep. 25, 2017), [http://www.slate.com/articles/technology/future\\_tense/2017/09/the\\_privacy\\_implications\\_of\\_using\\_digital\\_habits\\_to\\_track\\_mental\\_health.html](http://www.slate.com/articles/technology/future_tense/2017/09/the_privacy_implications_of_using_digital_habits_to_track_mental_health.html).

<sup>20</sup> Virtual hearing of the NCVHS Privacy, Confidentiality, and Security Subcommittee, (Nov. 28, 2017), <https://www.ncvhs.hhs.gov/transcripts-minutes/transcript-of-the-november-28-2017-meeting-of-the-privacy-confidentiality-and-security-subcommittee/>. See also Leslie P. Francis & John G. Francis, Data Re-Use and the Problem of Group Identity, 73 *Studies in Law, Politics, and Society* 143 (2017), <https://utah.pure.elsevier.com/en/publications/data-reuse-and-the-problem-of-group-identity>.

the clinical setting to help address a health concern.”<sup>21</sup> PGHD includes, but is not limited to health history, treatment history, biometric data, symptoms, and lifestyle choices.

PGHD is distinct from data generated in clinical settings and through encounters with providers in two important ways. First, patients, not providers, are primarily responsible for capturing or recording these data. Second, patients decide how to share or distribute these data to health care providers and others.<sup>22</sup>

New technologies enable patients to generate data outside of clinical settings and share it with providers. Examples of PGHD sources include blood glucose monitoring or blood pressure readings using home health equipment, and exercise and diet tracking using a mobile app. Smart phones, mobile applications and remote monitoring devices, when linked to the deployment of electronic health records (EHRs), patient portals, and secure messaging, will connect patients and providers.

Despite the considerable interest in PGHD, the capture, use, and sharing of PGHD for clinical care and research are not yet widespread. There are technical, legal, and administrative barriers, and the multiple stakeholders involved add more complexity. With continuing attention, improved technology, and increasing interoperability, these barriers are likely to be addressed by clinicians and researchers.

As HHS’s recent Non-Covered Entity Report illustrates, the privacy and security protections that apply to personal health records (PHRs) are uneven and may not be subject to a consistent legal and regulatory framework.<sup>23</sup> The same concerns about data integrity, security breaches, malware, and privacy identified for PHRs are likely to apply to PGHD. PGHD may originate with devices subject to HIPAA privacy and security rules from origin to destination and the data may originate with devices not subject to HIPAA or any other privacy or security requirements. PGHD that originates with unregulated commercial devices or that passes unencrypted over networks may be captured by third parties and used for consumer profiling or other purposes. Patients are likely to be in possession of copies of data that others maintain, and they may use or share the data as they please and without formal privacy protections in most instances. It may be challenging at times to tell whether and how PGHD falls under any definition of *health information*. It is possible to structure some device data activities so that the data falls under HIPAA at times, does not fall under HIPAA at times, and falls under HIPAA in the hands of

---

<sup>21</sup> Accenture, Conceptualizing a Data Infrastructure for the Capture, Use, and Sharing of Patient-Generated Health Data in Care Delivery and Research through 2024 at 2, (Draft White Paper for a PGHD Policy Framework) (Oct. 2016), [https://www.healthit.gov/sites/default/files/Draft\\_White\\_Paper\\_PGHD\\_Policy\\_Framework.pdf](https://www.healthit.gov/sites/default/files/Draft_White_Paper_PGHD_Policy_Framework.pdf). There are other definitions for PGHD, but the differences are not material here. Other relevant reports on PGHD include National eHealth Collaborative, Patient-Generated Health Information Technical Expert Panel (Dec. 2013), [https://www.healthit.gov/sites/default/files/pghi\\_tep\\_finalreport121713.pdf](https://www.healthit.gov/sites/default/files/pghi_tep_finalreport121713.pdf); Office of Policy and Planning, Office of the National Coordinator for Health Information Technology, Patient-Generated Health Data (Apr. 2012), [https://www.healthit.gov/sites/default/files/rti\\_pghd\\_whitepaper\\_april\\_2012.pdf](https://www.healthit.gov/sites/default/files/rti_pghd_whitepaper_april_2012.pdf).

<sup>22</sup> Id. at 33.

<sup>23</sup> Maximus Federal Services, Non-HIPAA Covered Entities: Privacy and Security Policies and Practices of PHR Vendors and Related Entities Report (2012) (prepared for the Office of the Chief Privacy Officer, Office of the National Coordinator for Health Information Technology), [https://www.healthit.gov/sites/default/files/maximus\\_report\\_012816.pdf](https://www.healthit.gov/sites/default/files/maximus_report_012816.pdf).

some participants but not in the hands of others. Data may move back and forth between different regulated and non-regulated regimes.

At the November 28, 2017 virtual hearing of the NCVHS Privacy, Confidentiality, and Security Subcommittee, Adam Greene, Partner at the law firm Davis Wright Tremaine, discussed the range of other law that might apply to non-HIPAA data and some of the shortcomings of those laws:

Just because HIPAA does not apply, though, does not mean that the information is unprotected under law. Information that HIPAA does not govern may be subject to the FTC Act, to state medical records laws, or to state consumer protection laws. The most significant challenge is that these laws usually offer little guidance in the area of information security, and may even include conflicting requirements.<sup>24</sup>

## 2. Trail of Data

Below is an example of basic activities that create data that most individuals would consider as health information, at least in an informal sense. While reading, keep in mind the different holders of data and the presence or absence of applicable privacy rules. Another factor is what standard might be used to tell if a particular type of data is health information. Outside of the formal definition as found in HIPAA, it may be largely a matter of personal judgment.

A mother finds her child has a cold and keeps her home from school. She calls to tell the school that her daughter won't be coming because of the cold. The mother goes to a pharmacy in a supermarket to buy an OTC cough medicine, stopping to ask the pharmacist for a recommendation. The next day, she takes her daughter to see a pediatrician. The pediatrician writes a prescription for a drug suitable for a child. The mother uses a drug manufacturer's coupon to buy the drug at the pharmacy. Along the way, she tells her boss about her daughter's cold and that she will be working at home for a few days. She posts a note on her private Facebook page. She sees her neighbor and explains why she is home that day. She researches the drug on the Internet, using a general search engine, a federal agency website, a Canadian website, and a commercial ad-supported medical information website where she is a registered user.

By following the data, the complexity of the definitional task beyond HIPAA becomes more apparent. The mother's activities produce some HIPAA-covered data, some data protected by other laws, and some unregulated data. This discussion is not fully complete, but it makes the point.

---

<sup>24</sup> Virtual hearing of the NCVHS Privacy, Confidentiality, and Security Subcommittee, (Nov. 28, 2017), <https://www.ncvhs.hhs.gov/transcripts-minutes/transcript-of-the-november-28-2017-meeting-of-the-privacy-confidentiality-and-security-subcommittee/>.

**School:** Most schools fall under privacy rules from the Federal Educational Rights and Privacy Act (FERPA).<sup>25</sup> Records subject to FERPA are exempt from HIPAA.<sup>26</sup> Health-related data is generally treated as an education record, although there are some complexities that are not of immediate interest here. The information given by the mother to the school is an education record subject to a privacy rule.

**Pharmacy:** The pharmacy is a HIPAA covered entity, so even the informal advice about an OTC product is PHI. Filling the prescription creates PHI as well, as is information shared with and obtained from a health plan, a pharmacy benefit manager, and a health care clearinghouse.

**Supermarket:** The purchase of the OTC medicine does not create PHI in the hands of the supermarket. Since the mother used a frequent shopper card to make the purchase, the supermarket can identify the mother, link the product with her other purchases, and sell any of the information to third party marketers or others. In general, customer information held by the supermarket is not subject to any privacy regulation.<sup>27</sup>

**Pediatrician:** A pediatrician is highly likely to be a HIPAA covered entity. The encounter between the child and the doctor also creates records at a health plan, pharmacy benefit manager, and health care clearinghouse, all HIPAA covered entities or business associates.

**Drug manufacturer:** A drug manufacturer acquires the mother's personal information because the coupon requires it. The drug manufacturer is not subject to HIPAA or likely to any other privacy law. The transaction record may give the drug manufacturer information about the drug purchased; the patient; the time, date and location of purchase; and the insurance policy that covered part of the price. Some or all of this data may be *health information*.

**Facebook:** As Facebook member, the mother can control posted data in various ways, but her Facebook "friends" have no obligation to treat it as private. Facebook can use the data for creating a member profile, for targeting advertising, and in other ways. Facebook can use information posted about health or health activities in the same way as other personal information.

**Federal website:** A federal informational website is not likely to collect or retain any identifiable information from a search. If the website maintained any

---

<sup>25</sup> 20 U.S.C. § 1232g, <https://www.law.cornell.edu/uscode/text/20/1232g>, 34 C.F.R. Part 99, <https://www.ecfr.gov/cgi-bin/text-idx?SID=497891dcf79707161b2903cbfde23de7&mc=true&node=pt34.1.99&rgn=div5>.

<sup>26</sup> 45 C.F.R. § 164.103 (definition of *protected health information*).

<sup>27</sup> There is at least one state law restricting disclosure of frequent shopper records. The law only applies to food retailers. See Supermarket Club Cards, California Civ. Code §§ 1749.60 - 1749.66, [http://leginfo.legislature.ca.gov/faces/codes\\_displayText.xhtml?lawCode=CIV&division=3.&title=1.4B.&part=4.&chapter=&article=](http://leginfo.legislature.ca.gov/faces/codes_displayText.xhtml?lawCode=CIV&division=3.&title=1.4B.&part=4.&chapter=&article=).

personal information, the data would likely, but not certainly, fall under the Privacy Act of 1974 whether characterized as health information or not.

**Canadian website.** A private sector website in Canada is subject to the Personal Information Protection and Electronic Documents Act (PIPEDA), a general privacy law regulating the private sector.<sup>28</sup> If the website collected any personal information, the information would be subject to the Canadian privacy law.

**Commercial U.S. medical information website.** U.S. websites, whether they collected health information or otherwise, are generally not subject to a federal privacy law. A website can be held to any privacy policy posted on its site.<sup>29</sup> The website can collect, retain, make use of, and share in various ways any information revealed by a search made by a visitor. In the example, the mother using the website registered previously on the site so that the website knew her identity and could add the latest inquiries to her profile. However, even in the absence of registration, it is possible that a website can identify visitors through the use of advertising trackers, cookies, IP addresses, or other means.

**Internet tracking and advertising companies, and Internet service provider.** While none of the website activities involved a direct disclosure to a tracking or advertising company, the possibility of a disclosure on commercial sites is high. In general, it is difficult or impossible for web users to know who is following their activities on any given website or from website to website. Clicking on an ad may result in additional disclosures. For example, if a company seeks to advertise only on web pages that show information to individuals with high incomes, the company knows something about the income of anyone who clicked on the ad.

**Neighbor.** Generally, no privacy law applies to disclosure of personal information to a neighbor. Even if all agree that the information is health information, it is hard to see any practical consequence to designation of the data in this context. Even broadly applicable EU privacy rules do not reach household activities.

**Boss.** The disclosure to the mother's workplace manager may be no more detailed than to a neighbor, but her employer is subject to some rules with respect to the use and disclosure of health information. Whether the information is health information with respect to the mother and how workplace rules apply to a disclosure about a dependent are more complicated questions than can be pursued here.

---

<sup>28</sup> Personal Information Protection and Electronic Documents Act, S.C. 2000, c. 5 (Can.), available at <http://canlii.ca/t/52hmg>.

<sup>29</sup> See, e.g., Federal Trade Commission, Enforcing Privacy Promises, <https://www.ftc.gov/news-events/media-resources/protecting-consumer-privacy/enforcing-privacy-promises>.

**Internet search engine.** An Internet search engine can, if it chooses, keep a record of searches made by its customers. It can build a user profile from the searches, from webpage visits, ad clicks, commercial databases, and more. It can then use that profile to make decisions about what search results to provide, what ads to show, or how to treat the customer in other ways. Whether a piece of information is health information or not may not make a difference.

### 3. Tentative Conclusions about Definitions

It is difficult to define *health information* without a context. It is important to know who the record keeper is, what the purpose of the processing is, and what the rights of the data subject are, if any. The same information that appears to be health information in one context may not be health information in another. Some physical characteristics bearing on health status (height, weight, age, some disabilities) are observable from physical presence or from photographs or videos.

Data that appears unrelated to health may be used to infer items of health information to a level of statistical likelihood.

Not all record keepers who hold health information have relationships with data subjects or are known to data subjects. This is true for some health record keepers who obtain PHI from HIPAA covered entities as well as for other record keepers whose data does not come from a source covered by HIPAA. Neither data nor the relationship between record keepers and data subjects is static.

Accepting for the moment that HIPAA solves the definitional problem well enough for HIPAA covered entities, it is not apparent that extending HIPAA automatically to others who hold *health information* will work. HIPAA ties its definition of PHI to the status of covered entities. With other types of record keepers, the broad scope of the HIPAA definition is likely to present serious difficulties and conflicts. For example, a bank may acquire health information in connection with a loan, but the HIPAA rule of treating all covered entity information as PHI would not work in a bank context, where some records are subject to other laws.<sup>30</sup>

Further, HIPAA strikes balances appropriate for the health care treatment and payment environments. The same balances and the same authorities to use and disclose health information would not be appropriate for other health record keepers. For example, it would not be appropriate to expressly authorize a convenience store that sells OTC drugs to make nonconsensual disclosures for treatment purposes in the same way that HIPAA authorizes a health care provider to make treatment disclosures. Nor would it be appropriate to authorize the store to disclose its information for the numerous other purposes allowed by physicians and insurers under HIPAA.

---

<sup>30</sup> See National Committee on Vital and Health Statistics, Recommendations on the financial services industry and § 1179 of HIPAA (Sep., 16, 2015), <https://www.ncvhs.hhs.gov/recommendations-on-the-financial-services-industry-and-%C2%A7-1179-of-hipaa/>

Establishing different rules for different record keepers has its attractions, but it faces the prospect of having different rules for the same information depending on the record keeper. Some record keepers with multiple functions might be subject to more than one set of rules at the same time.<sup>31</sup> Further, the large number of non-HIPAA record keepers makes rule making especially challenging, with those who derive health information from non-health data presenting a particular challenge. Even identifying all non-HIPAA health record keepers would be a challenge. On the other hand, the U.S. approach to privacy is sectoral, and all of the challenges described in this paragraph occur with other types of records, including HIPAA records.

If it is at all reassuring, the definitional problem raised here is not unique to the U.S. In the EU, where data protection rules treat all health data as sensitive information,<sup>32</sup> the European data protection supervisor concluded in the context of mobile health that there is not always a clear distinction between the health data and other types of “well-being” information that does not qualify as health data.<sup>33</sup>

The general problem of defining health information is harder to solve in a statute or rule that extends beyond HIPAA or outside the defined health sector. However, the problem may be less difficult if addressed in guidelines, standards, codes of conduct, best practices and other types of “soft” standards. Under a looser definition, a tradeoff between precision and consistency may arise, but the result could provide guidance good enough for at least some applications.

#### **D. Health Data Ownership, Control, and Consent**

Debates about health privacy occasionally include discussions about the ownership of health data. Traditionally, a physician owned the paper health record containing patient data.<sup>34</sup> As health care practice grew more complex in environments characterized by third party payment and electronic health records, issues about ownership grew more both more complex and less relevant. Traditional notions of property no longer had much meaning in the context of health records.<sup>35</sup> Further, in the digital environment, the same information can more easily be in multiple places and controlled by multiple persons at the same time. App developers and others make a business by collecting and exploiting patient health information and by asserting data ownership, control, or both. The issues here are conceptually and technically messy, unclear,

---

<sup>31</sup> This happens with federal agencies that are HIPAA covered entities. The agencies are subject to the Privacy Act of 1974 as well. In this case, the agency must comply with both laws. See HHS, How does the HIPAA Privacy Rule affect my rights under the Federal Privacy Act? (FAQ 351), <https://www.hhs.gov/hipaa/for-individuals/faq/351/how-does-hipaa-affect-my-rights-under-the-federal-privacy-act/index.html>.

<sup>32</sup> See, e.g., European Union, General Data Protection Regulation [2016] OJ L119/1, at Article 9, <http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52012PC0011>.

<sup>33</sup> European Data Protection Supervisor, Mobile Health Reconciling technological innovation with data protection at para. 16 (Opinion 1/2015), [https://edps.europa.eu/sites/edp/files/publication/15-05-21\\_mhealth\\_en\\_0.pdf](https://edps.europa.eu/sites/edp/files/publication/15-05-21_mhealth_en_0.pdf).

<sup>34</sup> See Mark A. Hall & Kevin A. Schulman, Ownership of Medical Information, 301 JAMA 1282 (2009), <https://jamanetwork.com/journals/jama/article-abstract/183601>. See also, George Washington University Health and the Law Project, Who Owns Medical Records: 50 State Comparison (2015), <http://www.healthinfolaw.org/comparative-analysis/who-owns-medical-records-50-state-comparison>.

<sup>35</sup> In a 2010 Australian case, a judge dissected the parts of a health record for the applicability of copyright, finding a copyright interest in some parts of the record and not others. For a summary and discussion of the case, see Judith Mair, Who owns the information in the medical record? Copyright issues (2011), [http://www.himaa.org.au/members/journal/HIMJ\\_40\\_3\\_2011/HIMJ\\_40-3\\_Mair\\_Medical\\_record\\_copyright.pdf](http://www.himaa.org.au/members/journal/HIMJ_40_3_2011/HIMJ_40-3_Mair_Medical_record_copyright.pdf).

and largely unaddressed in law. Some continue to promote patient ownership even as the notion of ownership becomes seemingly less meaningful.

## 1. The Regulated World

HIPAA sets out the rights and responsibilities of covered entities without discussing ownership. Covered entities have obligations under HIPAA – as well as under other laws, standard business practices, and ethical standards – to maintain records of health care treatment and payment. Patients have a bundle of rights under HIPAA and other laws that include the ability to inspect, have a copy of, and propose amendments to health records. Ownership seems irrelevant to the exercise of these rights and responsibilities. An Internet post about the issue of ownership of EHRs says that “Ownership, then, puts people in the wrong mindset.”<sup>36</sup>

Yet the concept of ownership persists in some quarters. Articles about the use of blockchain technology in health care sometimes tout patient “ownership” as a benefit of the technology.<sup>37</sup> It is not always clear from these references just what ownership means. In one proof-of-concept for blockchain in health care, “patients are enabled with fine-grained access control of their medical records, selecting essentially any portion of it they wish to share.”<sup>38</sup> It is pointless to debate whether control equates with ownership, but the notion of patient control of health records creates its own set of difficulties.

For health data regulated under HIPAA, a patient has limited ability to influence the way that covered entities use and disclose the patient’s health record. For most of the uses and disclosures allowed under HIPAA, a patient has virtually no say, and a covered entity can use and disclose PHI as allowed by the rule without seeking or obtaining patient consent. A patient can request restrictions on use or disclosure, but a covered entity need not consider or agree to any patient requests. A patient can influence some disclosures to care givers and for facility directories.

The lack of patient control is not necessarily a bad thing. Many activities in the health world require access to patient records, including payment of bills, health research, public health, oversight and accountability, and much more. Giving patients greater rights to consent – a position supported by some advocacy groups – requires the resolution of numerous conflicts between the public interest and a patient’s rights. HIPAA resolved those conflicts, although the choices could always be reopened. A health technology that gave patient greater ability to

<sup>36</sup> Wireless Life Sciences Alliance, Who Owns Your Health Care (EHR) Data?, <http://wirelesslifesciences.org/2015/09/who-owns-your-health-care-ehr-data/>.

<sup>37</sup> William Gordon, Adam Wright, & Adam Landman, Blockchain in Health Care: Decoding the Hype (2017), NEJM Catalyst <http://catalyst.nejm.org/decoding-blockchain-technology-health/> (“But perhaps the greatest potential of blockchain technology is the empowering of patients to own and gather their own data.”); RJ Krawiec, Dan Housman, Mark White, Mariya Filipova, Florian Quarre, Dan Barr, Allen Nesbitt, Kate Fedosova, Jason Killmeyer, Adam Israel, Lindsay Tsai, Blockchain: Opportunities for Health Care (2016) (Deloitte), <https://www2.deloitte.com/content/dam/Deloitte/us/Documents/public-sector/us-blockchain-opportunities-for-health-care.pdf>, (“Prescript, a proof-of-concept developed by Deloitte Netherlands, in collaboration with SNS Bank and Radboud, gives patients complete ownership of their medical records, allowing them to grant and revoke provider access to their data.”).

<sup>38</sup> Ariel Ekblaw, Asaph Azaria, John D. Halamka, Andrew Lippman, A Case Study for Blockchain in Healthcare: “MedRec” prototype for electronic health records and medical research data, MIT Media Lab (2016), <https://www.media.mit.edu/publications/medrec-whitepaper/>.

control use of their record would have to confront and resolve the conflicts. For example, if a patient could keep anyone from learning of a narcotics prescription, then it would be more difficult or impossible to prevent over-prescribing.

Before HIPAA, patients typically “controlled” the use and disclosure of their health records by signing “informed” consent forms presented to them by a health care provider. The forms were often not accompanied by any notice, and the authorization typically allowed the disclosure of “any and all” information to insurers and others. In the absence of a state law, there were no controls over uses by those who received the information. Patients signed the forms presented to them, with little opportunity to make any change. This old model produced a consent that was neither informed nor actually consensual.<sup>39</sup> Treatment and payment usually depended on a signature on the consent form.

HIPAA resolved the consent issue for the most part, but other paths seem possible. The technology that supports EHRs could support a greater role for patients. The promotion of blockchain as a tool for patient control is an example, but it does not appear that anyone has explored the limits of or mechanism for patient control in any depth. A recent article proposes that patients have a health data manager for a digital health record, with the relationship between patient and manager controlled through a data use agreement.<sup>40</sup> It is an outline of an idea for a greater role for the patient in some uses and disclosures. A defined role for patients would have to resolve at a societal level just what voice patients should have in disclosures for research, law enforcement, national security, health oversight, protection of the President, and more. Resolving those choices would then allow for designing a mechanism that patients could practically use. There appear to be many opportunities for exploring increased roles of patients in the use and disclosure of their HIPAA-regulated records. Technology can support more patient choice than was practical in the past.

## 2. The Unregulated World

So far, this discussion is mostly about consent and control in the regulated health care world. In the non-regulated world of health data, some record keepers have relationships with consumers (a fitness tracker provider may require a user to accept terms of service and to retrieve data from the provider’s website). Some record keepers may have no relationship with consumers. Data collected by websites and sold to data profilers, marketers, and others may have terms hidden from consumers or that give consumers no rights or interests in the data. Consumers may not even be aware of the extent of data collection, its use, or the identity of those in possession of the data.

An example of the complexity of relationships comes from Mindbody, a company that provides a technology platform for the wellness services industry.<sup>41</sup> The company’s role in wellness

---

<sup>39</sup> Robert Gellman, The Privacy of Health Information and the Challenge for Data Protection, Paper presented at Eighth International Conference of the Observatory "Giordano Dell'Amore" on the Relations Between Law and Economics, Stresa, Italy (May 1997), <https://www.bobgellman.com/rg-docs/rg-health-consent-97.pdf>.

<sup>40</sup> Katherine A. Mikk, Harry A. Sleeper & Eric J Topol, The Pathway to Patient Data Ownership and Better Health, 318 JAMA 1433 (2017), <https://jamanetwork.com/journals/jama/article-abstract/2654934>.

<sup>41</sup> <https://www.mindbodyonline.com/company>. The company provides both HIPAA and non-HIPAA covered services. The discussion here addresses services not covered by HIPAA.

activities may not be visible to individuals enrolled in a wellness service. The company's terms of service address data ownership and use by stating that the business providing the data owns the data as between the business and Mindbody. The policy does not address the rights of data subjects. However, the terms go on to give Mindbody broad rights to use and disclose the data:

You hereby grant to MINDBODY a nonexclusive, worldwide, assignable, sublicensable, fully paid-up and royalty-free license and right to copy, distribute, display and perform, publish, prepare derivative works of and otherwise use Your Data for the purposes of providing, improving and developing MINDBODY's products and services and/or complementary products and services of our partners.<sup>42</sup>

The data involved here has at least three parties in interest. There is the data subject who participates in the wellness program and provides some or all of the data in the program, the sponsor of the wellness program who contracts with Mindbody to provide technology services, and Mindbody itself. Mindbody's affiliates may represent another class of parties, as may those non-affiliated parties to whom Mindbody distributes data in accordance with the terms of service. The data subject here likely has little knowledge about most of the actual and potential data users resulting from enrollment in a wellness program.

Given all of the complexity of relationships involved in this one company's activities, the challenge of formally defining the rights and responsibilities of all the parties is daunting. Similar issues with multiple parties playing different roles and having different rights with respect to personally identifiable health data arise with devices like wearables and fitness monitors that monitor and record health information; Internet of Things devices that monitor eating, movement, and activities; personal health devices like a blood pressure cuff, and more.

The reaction of consumers when presented with the opportunity to give consent varies considerably. A recent research article contrasts patient responses to the collection and use of health data for unregulated activities with patient reaction for health research. Perhaps surprisingly, patients seemed indifferent to sharing with unregulated commercial entities, but patients showed "significant resistance" to scientific research applications.

Members of the general public expressed little concern about sharing health data with the companies that sold the devices or apps they used, and indicated that they rarely read the "terms and conditions" detailing how their data may be exploited by the company or third-party affiliates before consenting to them. In contrast, interviews with researchers revealed significant resistance among potential

---

<sup>42</sup> Mindbody Terms of Service, Data Ownership and Use, <https://www.mindbodyonline.com/terms-of-service#data-ownership-and-use>. Mindbody has a privacy policy that covers, among others, customers of subscribers. The policy reserves to Mindbody the right to make disclosures to a variety of third parties (including affiliates and non-affiliates) and gives individuals the right to opt-out of some (disclosures to non-affiliates) of those disclosures. Mindbody Privacy Policy (Oct. 18, 2016), <https://www.mindbodyonline.com/privacy-policy>.

research participants to sharing their user-generated health data for purposes of scientific study.<sup>43</sup>

Joseph Turow and his colleagues offer an explanation for the apparent lack of consumer interest in seeking to control online and other uses of their data. They find Americans resigned to the lack of control over data and powerless to stop its exploitation.

The findings also suggest, in contrast to other academics' claims, that Americans' willingness to provide personal information to marketers cannot be explained by the public's poor knowledge of the ins and outs of digital commerce. In fact, people who know more about ways marketers can use their personal information are more likely rather than less likely to accept discounts in exchange for data when presented with a real-life scenario.

Our findings, instead, support a new explanation: a majority of Americans are resigned to giving up their data—and that is why many appear to be engaging in tradeoffs. Resignation occurs when a person believes an undesirable outcome is inevitable and feels powerless to stop it. Rather than feeling able to make choices, Americans believe it is futile to manage what companies can learn about them. Our study reveals that more than half do not want to lose control over their information but also believe this loss of control has already happened.<sup>44</sup>

Whatever difficulties there are in giving consumers a greater say in the use of their regulated health data, the difficulties in doing the same in the unregulated world seem greater. In both worlds, narrow concepts of ownership of data or control of data do not appear helpful. The regulated world already has defined rights and responsibilities with health providers subject to ethical limitations on their actions. Yet ethical limits may not apply to other covered entities, including insurers and clearinghouses. Increasing consumer rights in the unregulated world may be harder because of the definitional difficulties already discussed; the likely and strong resistance of those profiting from the largely unrestricted use of health information; and a lack of political will. Changing the existing rules for the regulated world would be hard, but creating entirely new rules for the unregulated world would be even harder.

#### D. Fair Information Practices

Fair Information Practices (FIPs) are a set of internationally recognized practices for addressing the privacy of information about individuals. FIPs are important because they provide the underlying policy for many national laws addressing privacy and data protection matters. The international policy convergence around FIPs as core elements for information privacy has remained in place since the late 1970s. Privacy laws in the United States, which are much less

---

<sup>43</sup> Kirsten Ostherr, Svetlana Borodina, Rachel Conrad Bracken, Charles Lotterman, Eliot Storer and Brandon Williams, Trust and privacy in the context of user-generated health data, 4 *Big Data & Society* 1 (2017), <http://journals.sagepub.com/doi/full/10.1177/2053951717704673>.

<sup>44</sup> Joseph Turow, Michael Hennessy, & Nora Draper, THE TRADEOFF FALLACY How Marketers Are Misrepresenting American Consumers And Opening Them Up to Exploitation at 3(2015), [https://www.asc.upenn.edu/sites/default/files/TradeoffFallacy\\_1.pdf](https://www.asc.upenn.edu/sites/default/files/TradeoffFallacy_1.pdf).

comprehensive in scope than laws in some other countries, often reflect some elements of FIPs but not as consistently as the laws of most other nations.

FIPs are useful in understanding the elements of information privacy. HHS built the HIPAA privacy rule on a FIPs framework.

This final rule establishes, for the first time, a set of basic national privacy standards and fair information practices that provides all Americans with a basic level of protection and peace of mind that is essential to their full participation in their care.<sup>45</sup>

FIPs are a set of high-level policies and are not self-executing. Applying FIPs in any given context requires judgment rather than a mechanical translation. There can be disagreement on the best way to implement FIPs.

While it is fair to say that FIPs apply to covered entities through HIPAA, the application of FIPs to unregulated health information processors is less clear. Certainly FIPs can be applied if the will is there. Existing privacy policies and practices for these other activities are highly variable and only occasionally subject to any statutory standards. Anyone looking to devise a set of privacy policies for non-HIPAA health information activities might well choose to begin with FIPs. However, at the same time there is more universal recognition of the value of FIPs, some in the business community still would be happier if FIPs were edited to leave out standards they see as “inconvenient.”

While not of immediate relevance here, any health information activities in the European Union (EU) are subject to EU data protection law and to the FIPs principles embedded in that law. For example, any company that wants to sell fitness devices in the EU and to process the data that results would follow EU data protection law. That same company can sell the same device in the U.S. without any similar privacy protections.

---

<sup>45</sup> Department of Health and Human Services, Final Rule, *Standards for Privacy of Individually Identifiable Health Information*, 65 Federal Register 82462, 82464 (Dec. 28, 2000) at <http://www.gpo.gov/fdsys/pkg/FR-2000-12-28/pdf/00-32678.pdf>. See also id. at 82487 (“...our privacy regulation [is] based on common principles of fair information practices.”).

Table I: A Code of Fair Information Practices

1) The Principle of *Openness*, which provides that the existence of record-keeping systems and databanks containing data about individuals be publicly known, along with a description of main purpose and uses of the data.

2) The Principle of *Individual Participation*, which provides that each individual should have a right to see any data about himself or herself and to correct or remove any data that is not timely, accurate, relevant, or complete.

3) The Principle of *Collection Limitation*, which provides that there should be limits to the collection of personal data, that data should be collected by lawful and fair means, and that data should be collected, where appropriate, with the knowledge or consent of the subject.

4) The Principle of *Data Quality*, which provides that personal data should be relevant to the purposes for which they are to be used, and should be accurate, complete, and timely.

5) The Principle of *Use Limitation*, which provides that there must be limits to the internal uses of personal data and that the data should be used only for the purposes specified at the time of collection.

6) The Principle of *Disclosure Limitation*, which provides that personal data should not be communicated externally without the consent of the data subject or other legal authority.

7) The Principle of *Security*, which provides that personal data should be protected by reasonable security safeguards against such risks as loss, unauthorized access, destruction, use, modification or disclosure.

8) The Principle of *Accountability*, which provides that record keepers should be accountable for complying with fair information practices.

## II. Big Data: Expanding Uses and Users

### A. Overview of Big Data

The benefits of data and big data are unquestioned.<sup>46</sup> This is true whether the particular environment is a regulated or unregulated one. Even skeptics acknowledge the value of big data:

Big Data will lead to important benefits. Whether applied to crises in medicine, in climate, in food safety, or in some other arena, Big Data techniques will lead to significant, new, life-enhancing (even life-saving) benefits that we would be ill advised to electively forego.<sup>47</sup>

A huge number of reports from many sources address the general promise of big data as well as the promise of big data for health. Many reports also acknowledge the downsides that could result from some uses of big data. An Obama White House report from the President's Council of Advisors on Science and Technology (PCAST) focused on big data and privacy, acknowledging the promise and other consequences from big data in the first paragraph:

The ubiquity of computing and electronic communication technologies has led to the exponential growth of data from both digital and analog sources. New capabilities to gather, analyze, disseminate, and preserve vast quantities of data raise new concerns about the nature of privacy and the means by which individual privacy might be compromised or protected.<sup>48</sup>

The PCAST report also contained an observation on the data quality issues that may be an overlooked aspect of real-world and that may undermine the benefits of the data.

Real-world data are incomplete and noisy. These data-quality issues lower the performance of data-mining algorithms and obscure outputs. When economics allow, careful screening and preparation of the input data can improve the quality of results, but this data preparation is often labor intensive and expensive. Users, especially in the commercial sector, must trade off cost and accuracy, sometimes with negative consequences for the individual represented in the data. Additionally, real-world data can contain extreme events or outliers. Outliers may be real events that, by chance, are overrepresented in the data; or they may be the result of data-entry or data-transmission errors. In both cases they can skew the model and degrade performance. The study of outliers is an important research area of statistics.<sup>49</sup>

---

<sup>46</sup> See, e.g., Executive Office of the President, BIG DATA: SEIZING OPPORTUNITIES, PRESERVING VALUES (2014 Obama),

[https://obamawhitehouse.archives.gov/sites/default/files/docs/big\\_data\\_privacy\\_report\\_may\\_1\\_2014.pdf](https://obamawhitehouse.archives.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf).

<sup>47</sup> Paul Ohm, The Underwhelming Benefits of Big Data, 161 Univ of Pennsylvania L. Rev. 339(2013), [http://scholarship.law.upenn.edu/penn\\_law\\_review\\_online/vol161/iss1/22/](http://scholarship.law.upenn.edu/penn_law_review_online/vol161/iss1/22/).

<sup>48</sup> President's Council of Advisors on Science and Technology, Big Data and Privacy: A Technological Perspective at page ix (Obama 2014),

[https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast\\_big\\_data\\_and\\_privacy\\_may\\_2014.pdf](https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_big_data_and_privacy_may_2014.pdf).

<sup>49</sup> Id. at 25.

A Federal Trade Commission report also acknowledged the downsides of big data along with the benefits:

The analysis of this data is often valuable to companies and to consumers, as it can guide the development of new products and services, predict the preferences of individuals, help tailor services and opportunities, and guide individualized marketing. At the same time, advocates, academics, and others have raised concerns about whether certain uses of big data analytics may harm consumers, particularly low income and underserved populations.<sup>50</sup>

The FTC also summarized some of the benefits for health care:

Provide healthcare tailored to individual patients' characteristics. Organizations have used big data to predict life expectancy, genetic predisposition to disease, likelihood of hospital readmission, and likelihood of adherence to a treatment plan in order to tailor medical treatment to an individual's characteristics. This, in turn, has helped healthcare providers avoid one-size-fits-all treatments and lower overall healthcare costs by reducing readmissions. Ultimately, data sets with richer and more complete data should allow medical practitioners more effectively to perform "precision medicine," an approach for disease treatment and prevention that considers individual variability in genes, environment, and lifestyle.<sup>51</sup>

A 2013 report from McKinsey took a more focused look at health care use of big data. It recognized the need to protect privacy, but it suggested a need to shift the collective mind-set about patient data from protect to share with protections.<sup>52</sup> There are many data users, including researchers, who support greater access to data to be used for socially beneficial purposes. HIPAA supports research (and other) activities by making health data available for IRB-approved research without the need for patient consent.<sup>53</sup>

The White House PCAST report cited above offers a more specific example of the use of unregulated big data from sources such as cell phones, the Internet and personal devices to draw health conclusions:

Many baby boomers wonder how they might detect Alzheimer's disease in themselves. What would be better to observe their behavior than the mobile device that connects them to a personal assistant in the cloud (e.g., Siri or OK Google), helps them navigate, reminds them what words mean, remembers to do

---

<sup>50</sup> Federal Trade Commission, *BIG DATA A Tool for Inclusion or Exclusion? Understanding the Issues* (2016) at i, <https://www.ftc.gov/system/files/documents/reports/big-data-tool-inclusion-or-exclusion-understanding-issues/160106big-data-rpt.pdf>.

<sup>51</sup> *Id.* at 7.

<sup>52</sup> McKinsey&Company, *The 'big data' revolution in healthcare* (2013) at 13, [http://www.mckinsey.com/~media/McKinsey/dotcom/client\\_service/Healthcare%20Systems%20and%20Services/PDFs/The\\_big\\_data\\_revolution\\_in\\_healthcare.ashx](http://www.mckinsey.com/~media/McKinsey/dotcom/client_service/Healthcare%20Systems%20and%20Services/PDFs/The_big_data_revolution_in_healthcare.ashx).

<sup>53</sup> 45 C.F.R. § 164.512(i).

things, recalls conversations, measures gait, and otherwise is in a position to detect gradual declines on traditional and novel medical indicators that might be imperceptible even to their spouses?

At the same time, any leak of such information would be a damaging betrayal of trust. What are individuals' protections against such risks? Can the inferred information about individuals' health be sold, without additional consent, to third parties (e.g., pharmaceutical companies)? What if this is a stated condition of use of the app? Should information go to individuals' personal physicians with their initial consent but not a subsequent confirmation?<sup>54</sup>

This example raises direct questions about proper uses for unregulated health data and the likely lack of any patient protections for the data in the hands of cell phone providers, apps developers, and others. Frank Pasquale, Professor of Law, University of Maryland, discussed in a recent article the vast scope of data broker records, the potential uses for those records, and lack of auditing of for potentially illegal applications of unregulated health data:

While the intricate details of the Omnibus HIPAA rule are specified and litigated, data brokers continue gathering information, and making predictions based on it, entirely outside the HIPAA-protected zone. It is increasingly difficult for those affected to understand (let alone prove) how health-inflected data affected decision-making about them, but we now know that health-based scoring models are common. While the sheer amount of data gathered by reputational intermediaries is immense, the inferences they enable are even more staggering. Unattributed data sources are available to be pervasively deployed to make (or rationalize) critical judgments about individuals. Even if some of those judgments violate the law, there is no systematic auditing of data used by large employers in their decision-making, and there are ample pretexts to mask suspect or illegal behavior.<sup>55</sup>

The combination of vast amounts of data; use of advanced algorithms and artificial intelligence; and lack of both regulation and oversight lead Professor Pasquale to say that for health data outside the healthcare sector, "in many respects, it is anything goes."<sup>56</sup>

The amount of PII collected and available increases with new technologies.<sup>57</sup> Surveillance is pervasive, with cameras and automated license plate readers commonplace today. Tracking of

---

<sup>54</sup> President's Council of Advisors on Science and Technology, Big Data and Privacy: A Technological Perspective at 13-14 (Obama 2014),

[https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast\\_big\\_data\\_and\\_privacy\\_-\\_may\\_2014.pdf](https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_big_data_and_privacy_-_may_2014.pdf).

<sup>55</sup> Frank Pasquale, Redescribing Health Privacy: The Importance of Information Policy, 14 Hous. J. Health L. & Policy 95, 108 (2014) (footnotes omitted), [https://www.law.uh.edu/hjhlp/volumes/Vol\\_14/Pasquale.pdf](https://www.law.uh.edu/hjhlp/volumes/Vol_14/Pasquale.pdf).

<sup>56</sup> Virtual hearing of the NCVHS Privacy, Confidentiality, and Security Subcommittee, (Nov. 28, 2017) <https://www.ncvhs.hhs.gov/transcripts-minutes/transcript-of-the-november-28-2017-meeting-of-the-privacy-confidentiality-and-security-subcommittee/>.

<sup>57</sup> See generally Government Accountability Office, Information Resellers, Consumer Privacy Framework Needs to Reflect Changes in Technology and the Marketplace (2013), <http://www.gao.gov/assets/660/658151.pdf>.

Internet activity is a major industry. Linking of consumer devices that allows tracking of an individual's activity whether on a computer, cell phone, or other device is routine.<sup>58</sup>

Frequent shopper cards used at supermarkets and other merchants produce details accounts of consumer purchases. So does Internet shopping. Companies manage huge databases about consumers, with one company, for example, claiming it has over 800 billion consumer attributes.<sup>59</sup> How much of this data is or could be health information returns to the issue of what is health data.

It is useful to ground this discussion with a real world example. LexisNexis Risk Solutions, a large data broker and analytics company “health risk prediction scores independent of traditional health care data that leverage hundreds of key attributes found within public records data to provide health care entities with a picture of unforeseen and avoidable risks.”<sup>60</sup> The product brochure says on the cover “Predict health risk more precisely—without medical claims data.”<sup>61</sup> The result is something that LexisNexis promotes as health data but that relies on no actual PHI derived from HIPAA regulated records.

Companies and governments can use results of scores and other algorithmic products in many ways, with opaque decision making processes that can affect individuals without their being aware.

Eligibility decisions for a loan, other financial services, insurance, healthcare, housing, education, or employment can have significant and immediate impacts by excluding people outright. Equally significant economic effects can stem from less favorable service, terms, or prices, such as through fees, interest rates, or insurance premiums. Data driven decision may either occur in a fully automated manner, as in the case of a bank account or credit application denial, or they may happen prior to the actual decision, for example, when unwanted people are automatically rated low and filtered out, and thus never seen by human staff or by a system farther on in the process.<sup>62</sup>

Big data and the algorithms that use big data can produce information that looks like health information and can be used as health information but that has no actual health content. Further, this information can be used to make determination about individuals without the knowledge of

---

<sup>58</sup> The Federal Trade Commission conducted a workshop on cross-device tracking and subsequently issues a staff report. Cross-Device Tracking: A Federal Trade Commission Staff Report (2017), [https://www.ftc.gov/system/files/documents/reports/cross-device-tracking-federal-trade-commission-staff-report-january-2017/ftc\\_cross-device\\_tracking\\_report\\_1-23-17.pdf](https://www.ftc.gov/system/files/documents/reports/cross-device-tracking-federal-trade-commission-staff-report-january-2017/ftc_cross-device_tracking_report_1-23-17.pdf)

<sup>59</sup> Verisum, <https://versium.com/lifedata-matching-technology-can-help-fix-data-silos/>.

<sup>60</sup> LexisNexis Risk Solutions, LexisNexis Socioeconomic Health Scores, <https://www.lexisnexis.com/risk/downloads/literature/health-care/Socioeconomic-Health-Risk-Score-br.pdf>.

<sup>61</sup> Id.

<sup>62</sup> Wolfie Christl, HOW COMPANIES USE PERSONAL DATA AGAINST PEOPLE Automated Disadvantage, Personalized Persuasion, and the Societal Ramifications of the Commercial Use of Personal Information at 17-18 (2017) (Cracked Labs), [https://digitalcourage.de/sites/default/files/users/161/crackedlabs\\_christl\\_dataagainstpeople.pdf](https://digitalcourage.de/sites/default/files/users/161/crackedlabs_christl_dataagainstpeople.pdf).

the data subject and even without any direct human involvement. The section of this report that addresses other laws pursues this issue in the discussion of the Fair Credit Reporting Act.

## B. Defining Big Data

No serious person doubts the value of data in human endeavors, and especially in science, health care, public policy, education, business, history, and many other activities.<sup>63</sup> When *data* becomes *big data* is not entirely clear.<sup>64</sup> A common description is “[b]ig data is a term for data sets that are so large or complex that traditional data processing application software is inadequate to deal with them.”<sup>65</sup> Some definitions focus on the “high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making.”<sup>66</sup> Another definition adds veracity and value as fourth and fifth “v’s,” along with volume, velocity, and variety.<sup>67</sup> Perhaps vague is another word applicable to big data. Two scholars call big data “the buzzword of the decade.”<sup>68</sup>

There is no reason here to attempt a definition. However, it is noteworthy that a common feature of the definitions is a lack of any formal, objective, and clear distinction between data and big data. A statutory definition that draws a bright line is absent.<sup>69</sup> That presents a challenge for any regulation, a challenge similar to the problem of defining health information outside the HIPAA context. What cannot be defined cannot be regulated.

A good example of big data outside HIPAA regulation comes from the Precision Medicine Initiative, now known as the All of Us Research Initiative.<sup>70</sup> This program involves the building of a national research cohort of one million or more U.S. participants. The protocol for the initiative describes the plan for the dataset the program will maintain.

Ideally, in time, the core dataset will include participant provided information (PPI), physical measurements, baseline biospecimen assays, and baseline health

---

<sup>63</sup> The issues raised here are international in scope. See, e.g., Rosemary Wyber, Samuel Vaillancourt, William Perry, Priya Mannava, Temitope Folaranmi & Leo Anthony Celi, Big data in global health: improving health in low- and middle-income countries, 93 *Bulletin of the World Health Organization* 203 (2015), <http://www.who.int/bulletin/volumes/93/3/14-139022/en/>. This short but useful article is a summary of many of the issues identified here.

<sup>64</sup> See, e.g., MIT Technology Review, *The Big Data Conundrum: How to Define It?* (Oct. 2013), <https://www.technologyreview.com/s/519851/the-big-data-conundrum-how-to-define-it/>.

<sup>65</sup> Big Data entry (Oct. 9, 2017), Wikipedia, [https://en.wikipedia.org/wiki/Big\\_data](https://en.wikipedia.org/wiki/Big_data).

<sup>66</sup> See, e.g., Gartner, Inc., *IT Glossary*, cited in President’s Council of Advisors on Science and Technology, *Big Data and Privacy: A Technological Perspective* (Obama 2014), [https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast\\_big\\_data\\_and\\_privacy\\_-\\_may\\_2014.pdf](https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_big_data_and_privacy_-_may_2014.pdf).

<sup>67</sup> See, e.g., Bernard Marr, *Why only one of the 5 Vs of big data really matters* (IBM Big Data & Analytics Hub 2015), <http://www.ibmbigdatahub.com/blog/why-only-one-5-vs-big-data-really-matters>.

<sup>68</sup> Solon Barocas and Andrew D. Selbst, *Big Data’s Disparate Impact*, 104 *Calif. L. Rev.* 671 (2016), <http://www.californialawreview.org/wp-content/uploads/2016/06/2Barocas-Selbst.pdf>.

<sup>69</sup> None of the eight bills introduced through October 9, 2017, that contain the words *big data* offers a definition.

<sup>70</sup> See <https://www.nih.gov/AllofUs-research-program/pmi-cohort-program-announces-new-name-all-us-research-program>. See also testimony of Stephanie Devaney, All of Us Research Program, National Institutes of Health, National Committee on Vital and Health Statistics Full Committee (Sep. 13, 2017), <https://www.ncvhs.hhs.gov/transcripts-minutes/transcript-of-the-september-13-2017-ncvhs-full-committee-meeting/>.

information derived from EHRs from most participants. Data elements will be transferred through encrypted channels to the core dataset, which will be stored in the All of Us Research Program Data and Research Center (DRC).<sup>71</sup>

The dataset will include information from each participant's EHR, and the data will be updated over time.<sup>72</sup> Individuals must consent to participate in the program, and the consent process will include educational materials to enable participants to understand the program.<sup>73</sup> The program has a detailed Data Security Policy Principles and Framework<sup>74</sup> and has a certificate of confidentiality that provides privacy protections suitable for research activities.<sup>75</sup>

The potential value of the All of Us Research Initiative is not in question here. However, a privacy advocacy group raised questions about the adequacy legal protections for the research dataset.<sup>76</sup> HIPAA does not apply to NIH or to the records in the dataset, nor does the Privacy Act of 1974.<sup>77</sup> The research protocol itself notes "the risk that a third party may ask the All of Us Research Program to disclose information about participants without their permission as part of legal or other claims." Those third parties could include law enforcement and national security agencies. Other third parties may be able to obtain All of Us records from participants with their consent. Insurance companies and employers are examples of third parties with some degree of power over consumers.

The All of Us Research Initiative is an example of a big data type of activity that results in the creation of a large dataset of health records outside the health care treatment and payment system. Legal and privacy protections for the program are not the same as the protections for HIPAA records. All of Us obtains records with the informed consent of data subjects, a process that results in records no longer subject to HIPAA protections in the hands of a recipient who is not otherwise a HIPAA covered entity. If a private company undertook a similar, consent-based, activity, it could establish its own health database outside HIPAA and subject only to its own policies.<sup>78</sup>

Nothing here should be read to suggest that the All of Us Research Initiative is ill-conceived, poorly intentioned, or lacks a privacy policy. The program's privacy policy and certificate of confidentiality provide significant protections for privacy, albeit not the same as traditional health records. However, as a big data resource, All of Us stands outside the statutory

---

<sup>71</sup> National Institutes of Health, All of Us Research Protocol at 10 (undated) (Core Protocol V1), [https://allofus.nih.gov/sites/default/files/allofus-initialprotocol-v1\\_0.pdf](https://allofus.nih.gov/sites/default/files/allofus-initialprotocol-v1_0.pdf).

<sup>72</sup> Id.

<sup>73</sup> Id. at 18.

<sup>74</sup> <https://allofus.nih.gov/sites/default/files/privacy-trust-principles.pdf> and <https://allofus.nih.gov/sites/default/files/security-principles-framework.pdf>.

<sup>75</sup> National Institutes of Health, All of Us Research Protocol at 39 (undated) (Core Protocol V1), [https://allofus.nih.gov/sites/default/files/allofus-initialprotocol-v1\\_0.pdf](https://allofus.nih.gov/sites/default/files/allofus-initialprotocol-v1_0.pdf).

<sup>76</sup> World Privacy Forum, Privacy, the Precision Medicine Initiative, & the All of Us Research Program: Will Any Legal Protections Apply? (2017), <https://www.worldprivacyforum.org/2017/03/report-privacy-the-precision-medicine-initiative-and-all-of-us-research-program-will-any-legal-protections-apply/>.

<sup>77</sup> 5 U.S.C. § 552a, <https://www.law.cornell.edu/uscode/text/5/552a>.

<sup>78</sup> See, e.g., Antonio Regalado, Google's Health Study Seeks 10,000 Volunteers to Give Up Their Medical Secrets, MIT Technology Review (Apr. 19, 2017), <https://www.technologyreview.com/s/604224/googles-massive-health-study-seeks-10000-volunteers-to-give-up-their-medical-secrets/>.

protections available for those traditional records. Data subject consent allows All of Us to proceed in this manner because consent effectively eliminates the protections that HIPAA imposes on health records held by HIPAA covered entities. Other traditional protections, like the physician-patient testimonial privilege, may be weakened for records disclosed with the consent of a patient.

Another example of a pool of health data maintained outside of HIPAA regulations comes from prescription drug monitoring programs (PDMP). PDMPs are state-run electronic databases that track the prescribing and dispensing of controlled prescription drugs. They give health care providers access to data about a patient's controlled substance prescription history. This allows providers to identify patients who at risk of misusing controlled substances. Of course, PDMPs also assist law enforcement.<sup>79</sup> While PDMPs are not typically thought of as a big data resource, the databases collectively contain large amounts personally identifiable health information not regulated by HIPAA because no covered entity maintains the data. Leo Beletsky, Associate Professor of Law and Health Sciences at Northwestern University recently said that while PDMPs are an essential tool in combating drug abuse, the data may be available to a "wide variety of actors" and that there are "a number of privacy issues with these programs that have not received adequate attention."<sup>80</sup>

PatientsLikeMe is different example of a pool of health data outside of HIPAA. It is a company that maintains "a free website where people can share their health data to track their progress, help others, and change medicine for good."<sup>81</sup> The company enrolls patients, collects their data, and sells it to partners, including companies that develop or sell products (e.g., drugs, devices, equipment, insurance, and medical services) to patients. The company does not sell personally identifiable information (PII) for marketing purposes.<sup>82</sup> It also shares data with fellow patients. The company has over 600,000 members, covers 2800 health conditions, published over 100 research studies, and maintains over 43 million data points.<sup>83</sup> In general, PatientsLikeMe offers a model for supporting health research that differs in substantial ways from traditional privacy-protecting policies (like HIPAA) and ethic policies (like the Common Rule).

A final example comes from the traditional commercial world of data brokers. Individually, no one list, data resource, or company may "qualify" to be categorized as big data (even though there is no definition for that term). Collectively, however, all the data brokers and, perhaps, some individual data brokers have sufficient volume of information to support a big data label.

The example selected here comes from a company named Complete Medical Lists.<sup>84</sup> This list offers information about Diabetes Sufferers.<sup>85</sup> It is just one of many similar lists the company

<sup>79</sup> See Interim Report, President's Commission on Combating Drug Addiction and the Opioid Crisis at page 6 (undated draft 2017), <https://www.whitehouse.gov/sites/whitehouse.gov/files/ondcp/commission-interim-report.pdf>.

<sup>80</sup> Greg Slabodkin, Health Data Management, Prescription drug monitoring programs come under fire (Oct. 27, 2017), <https://www.healthdatamanagement.com/news/prescription-drug-monitoring-programs-come-under-fire>.

<sup>81</sup> PatientsLikeMe, About Us, <https://www.patientslikeme.com/about>.

<sup>82</sup> PatientsLikeMe, How does PatientsLikeMe make money?, <https://support.patientslikeme.com/hc/en-us/articles/201245750-How-does-PatientsLikeMe-make-money->.

<sup>83</sup> The company's website lists these numbers on its homepage. <https://www.patientslikeme.com/>.

<sup>84</sup> [http://completemedicallists.com/about\\_medical\\_lists.php](http://completemedicallists.com/about_medical_lists.php).

<sup>85</sup> [http://completemedicallists.com/mailling\\_lists.php?id=42&Ailments:%20Diabetes%20Sufferers](http://completemedicallists.com/mailling_lists.php?id=42&Ailments:%20Diabetes%20Sufferers).

(and the rest of the data broker industry) offers. The advertised number of individuals and their type of diabetes is:

973,771 Total Diabetes Sufferers	62,547 Total with Juvenile Diabetes
300,024 Total with Diabetes Type 1	508,364 Total with Diabetes Type 2
888,213 With Telephone Numbers	79,047 With Email Addresses

The company offers additional selections from its list based on the types of treatments used.

104,080 Avandia	163,785 Glucophage
108,459 Glucotrol	196,370 Insulin
48,939 Insulin Pump	5,324 Insulin - Lantus
164,970 Metformin HCl	86,899 Oral Medication
67,005 Other	153,027 Actos
7,220 Insulin - 1 or 2 times per day	5,792 Insulin - 3+ times per day
1,753 Insulin – Humulin	1,395 Insulin - Novolin
28,681 Uses Oral Medication	43,483 Insulin Injection

Also available are other selections based on non-health characteristics.

Phone Number	Age
Income	Gender
Presence of Children	Marital Status
Education	Occupation
Credit Card	Homeowner
Geography	Key Code
Hotline <sup>86</sup>	

The company does not identify sources of the health data for this list, but it is probably safe to assume that the data does not originate with any HIPAA covered entity. Some of the data may come directly from patients who fill out commercial surveys, register at websites, or perhaps from other tracking of Internet activities. Much of the data broker industry is unknown to most consumers. The company's website, which offers a product aimed at commercial users, does not appear to have a privacy policy.

### C. Big Data and Privacy

Big data presents conflicts with core privacy values. Some of the harder conflicts arise over the core Fair Information Practices (FIPs) principles of collection limits and purpose specification. The discussion that follows seeks to provide a flavor of the ongoing debates. A full treatment of the issues would exceed the scope of this report.

A core privacy principle states that there should be limits to the collection of personal data and any such data should be obtained by lawful and fair means and, where appropriate, with the knowledge or consent of the data subject. Some proponents of big data argue that it may not

<sup>86</sup> The datacard for this offering has a date of 9/28/2017.

always be possible to determine in advance whether data would be valuable so that it would be counterproductive to limit data collection.

For example, the PCAST report discusses the difficulty of protecting privacy by controlling the collection of privacy-sensitive data.

It is also true, however, that privacy-sensitive data cannot always be reliably recognized when they are first collected, because the privacy-sensitive elements may be only latent in the data, made visible only by analytics (including those not yet invented), or by fusion with other data sources (including those not yet known). Suppressing the collection of privacy-sensitive data would thus be increasingly difficult, and it would also be increasingly counterproductive, frustrating the development of big data's socially important and economic benefits.<sup>87</sup>

The slippery slope here is that if the value of data is not always knowable in advance, then all data should be collected and maintained without limit. The privacy community finds that argument unacceptable. It is not enough that the resulting all-inclusive record of personal activities, interests, locations, and more might someday product some useful insights. The political, social, personal, and economic consequences of comprehensive surveillance and recording would not be acceptable. A particular concern to many is the possibility that big data collected for research or for private sector would be used by government to make decisions about individuals (e.g., who can board an airplane, who can stay in the US, who retains eligibility for government programs, etc.). That is the heart of the tension over collection limits. Mark Rothstein observes a similar tension in the research arena, where some see promised big data capabilities as a reason to weaken research ethics rules. Rothstein rejects those pressures.<sup>88</sup>

The Article 29 Data Protection Working Party is an organization of national data protection authorities established under the EU Data Protection Directive.<sup>89</sup> The Article 29 Working Party issued a short opinion in 2014 about the conflicts between big data and privacy. In general, the opinion held the line and conceded little ground to big data pressures, noting that the challenges of big data might at most require “innovative thinking” on how key data protection principles apply in practice. However, the Working Party did not abandon or suggest weakening any privacy principles. In contrast, the Working Party observed that the “benefits to be derived from big data analysis can therefore be reached only under the condition that the corresponding

---

<sup>87</sup> President's Council of Advisors on Science and Technology, Big Data and Privacy: A Technological Perspective at page 47 (Obama 2014), [https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast\\_big\\_data\\_and\\_privacy\\_-\\_may\\_2014.pdf](https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_big_data_and_privacy_-_may_2014.pdf).

<sup>88</sup> Mark A. Rothstein, Ethical Issues in Big Data Health Research, 43 Journal of Law, Medicine, and Ethics page 425 (2015), [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2535373##](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2535373##), (“The development of new analytical tools, however, such as Big Data, should not serve as a catalyst for abandoning foundational principles of research ethics.”).

<sup>89</sup> Article 29 Data Protection Working Party, Statement on Statement of the WP29 on the impact of the development of big data on the protection of individuals with regard to the processing of their personal data in the EU (2014) (WP 221), [http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp221\\_en.pdf](http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp221_en.pdf).

privacy expectations of users are appropriately met and their data protection rights are respected.”

With respect to data collection, the Working Party observed that it “needs to be clear that the rules and principles are applicable to all processing operations, starting with collection in order to ensure a high level of data protection.” In other words, the Working Party rejected the argument that justifies the collection of personal data because the data might possibly be useful someday.

A second privacy principle (“purpose specification”) requires that the purposes for personal data collection should be specified not later than at the time of data collection and the subsequent use limited to the fulfillment of those purposes or such others as are not incompatible with those purposes. Some propose that the purpose specific standard is too limiting in light of the potential of big data, and that the standard should focus instead on the “interests that are served by the use of the collected data.”<sup>90</sup> A focus on uses and on the potential harm to individuals from those uses is a familiar argument from those looking for approaches to privacy different from the traditional Fair Information Practices.<sup>91</sup> A harm test raises a host of definitional problems of its own, and there is considerable ongoing litigation trying to define what harm is and when it deserves compensation. The Federal Trade Commission recently announced a workshop on information harms.<sup>92</sup> One of the problems with a harm standard is that it allows almost any use or disclosure unless the data subject can provide (in a court of law) that a direct economic harm resulted. That can be a very high barrier. A harm standard seems to deny a data subject any rights absent harm.

The Article 29 Working Party found the purpose specification principle still important to data protection.<sup>93</sup> It observed that “in particular, upholding the purpose limitation principle is essential to ensure that companies which have built monopolies or dominant positions before the development of big data technologies hold no undue advantage over newcomers to these markets.” The purpose limitation principle, together with the use limitation principle, set boundaries on how big (or other) data affects privacy interests. The intersection of data with algorithms, artificial intelligence, and analytics is an appropriate focal point for analysis, more so than just the data by itself. The studies cited and quoted here give much attention to the possibility that algorithms and similar tools can have deleterious consequences.

Interestingly, the Working Party viewed privacy as “essential to ensure fair and effective competition between economic players on the relevant markets.” It also noted that EU implementation of comprehensive information systems in the delivery of health services, among

---

<sup>90</sup> See, e.g., Lokke Moerel and Corien Prins, Privacy for the *homo digitalis* Proposal for a new regulatory framework for data protection in the light of Big Data and the Internet of Things at page 2 (2016), <http://ssrn.com/abstract=2784123>.

<sup>91</sup> See, e.g., Fred H. Cate, The Failure of Fair Information Practice Principles (2006), in Jane K Winn, *Consumer Protection in the Age of the Information Economy* (UK 2006), <https://ssrn.com/abstract=1156972>.

<sup>92</sup> Federal Trade Commission, Press Release, FTC Announces Workshop on Informational Injury, Sep. 29, 2016, <https://www.ftc.gov/news-events/press-releases/2017/09/ftc-announces-workshop-informational-injury>.

<sup>93</sup> See generally, Article 29 Data Protection Working Party, Opinion 03/2013 on purpose limitation (2013) (WP 203), [http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2013/wp203\\_en.pdf](http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf).

other large data systems, occurred under the traditional EU data protection standards. In other words, data protection rules did not interfere with EHRs and similar health sector information systems.

#### **D. Other Concerns about Big Data**

Big data presents both new and familiar types of threats to privacy or to other interests. Any compilation of new data, especially if the data is unregulated health data, may exacerbate existing privacy concerns. The willingness of patients to provide data to the health care system and to participate in research activities can be undermined if more individuals see big data – and especially more unregulated health data – as threatening the protections that exist today. There are many threats to privacy from expanded data activities and new technology, so big data is just another threat on this scale. Whether the benefits of big data are overhyped remains an open question.

One set of concerns is that the creation, collection, and maintenance of more data will undermine the ability to employ de-identified data as an alternative to sharing identifiable PII. Undermining the utility of de-identified data not only affects the privacy of data subjects, but it may make it harder to rely on de-identified data for research and other socially beneficial activities. For example, in a recent letter to the Secretary, NCVHS discussed risks of re-identification posed by longitudinal databases.<sup>94</sup> More data in more databases is another aspect of big data.

The threat of data re-identification arises whether the de-identified data at issue is HIPAA data or unregulated health data. New capabilities to re-identify data may ultimately require adjustments in the HIPAA standards. The unregulated world of data and health data has no standards to adjust.

Another concern has to do with the uses of big data. Put another way, the concern is not so much about the data itself but about the way that business, government, researchers, and others use the data. The focus here is on the analytics fueled by big data. A significant concern is that big data will support inequality and bias.

Approached without care, data mining can reproduce existing patterns of discrimination, inherit the prejudice of prior decision makers, or simply reflect the widespread biases that persist in society. It can even have the perverse result of exacerbating existing inequalities by suggesting that historically disadvantaged groups actually deserve less favorable treatment.<sup>95</sup>

The PCAST report made a similar point about the shortcomings of data analytics and the potential for biased and discriminatory consequences.

Many data analyses yield correlations that might or might not reflect causation. Some data analyses develop imperfect information, either because of limitations

<sup>94</sup> Recommendations on De-identification of Protected Health Information under HIPAA (Feb. 23, 2017), <https://www.ncvhs.hhs.gov/recommendations-on-de-identification-of-protected-health-information-under-hipaa/>.

<sup>95</sup> Solon Barocas and Andrew D. Selbst, Big Data's Disparate Impact, page 104 Calif. L. Rev. 671, 674 (2016), <http://www.californialawreview.org/wp-content/uploads/2016/06/2Barocas-Selbst.pdf>.

of the algorithms, or by the use of biased sampling. Indiscriminate use of these analyses may cause discrimination against individuals or a lack of fairness because of incorrect association with a particular group. In using data analyses, particular care must be taken to protect the privacy of children and other protected groups.<sup>96</sup>

This does not exhaust concerns about big data and privacy.<sup>97</sup> Professor Dennis Hirsch writes about the risks of predictive analytics, an application of big data that seeks correlations in large data sets to learn from past experience to predict the future behavior of individuals in order to drive better decisions. Predictive analytics have many positive and other applications. Hirsch usefully identifies four risks of predictive analytics. The first is a privacy risk. The analysis may identify information about an individual that the individual does not care to share (e.g., pregnancy, political views, sexual orientation). The second risk is a bias risk. Neutral application of analytics may result in discrimination against protected classes. A third risk is error risk. Incorrect or incomplete facts and flawed algorithms can lead to wrong predictions that harm people (e.g., preventing someone from boarding an airplane). The fourth risk is exploitation risk, which is taking advantage of vulnerable people (for example, building a list for use by scammers). This is a helpful framework for thinking about applications of big data and analytics.<sup>98</sup>

### **E. Responses to Big Data**

For HIPAA data, the Privacy Rule operates to control the flow and use of PHI in big and small quantities. The adequacy of the HIPAA rule is an assumption of this report. Looking just at health data not covered by HIPAA to find possible responses is not a simple task.

There is no consensus about the scope or existence of big data problems, and responses to those problems suffer from the same lack of agreement. There are, however, useful discussions in this arena. A review conducted for the Obama Administration commented on the risks and rewards of big data.

An important finding of this review is that while big data can be used for great social good, it can also be used in ways that perpetrate social harms or render outcomes that have inequitable impacts, even when discrimination is not intended. Small biases have the potential to become cumulative, affecting a wide range of outcomes for certain disadvantaged groups. Society must take steps to guard against these potential harms by ensuring power is appropriately balanced

---

<sup>96</sup> President's Council of Advisors on Science and Technology, Big Data and Privacy: A Technological Perspective at page 25 (Obama 2014), [https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast\\_big\\_data\\_and\\_privacy\\_-\\_may\\_2014.pdf](https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_big_data_and_privacy_-_may_2014.pdf).

<sup>97</sup> For a list of other privacy risks, see, Massachusetts Institute of Technology, Workshop Summary Report, Big Data Privacy Workshop Advancing the State of the Art in Technology and Practice (2015) at page 6, [http://web.mit.edu/bigdata-priv/images/MITBigDataPrivacyWorkshop2014\\_final05142014.pdf](http://web.mit.edu/bigdata-priv/images/MITBigDataPrivacyWorkshop2014_final05142014.pdf).

<sup>98</sup> Dennis Hirsch, Predictive Analytics Law and Policy: A New Field Emerges, *I/S: A Journal of Law and Policy for the Information Society* (2017) (forthcoming).

between individuals and institutions, whether between citizen and government, consumer and firm, or employee and business.<sup>99</sup>

Some protections or responses for specific applications may come from existing laws. For example, scholars examined remedies from Title VII of the Civil Rights Act<sup>100</sup> for discriminatory data mining and other activities affecting employment.<sup>101</sup> Where big data analytics activities result in civil rights violations, there may be remedies in existing laws that prohibit discrimination based on protected characteristics such as race, color, sex or gender, religion, age, disability status, national origin, marital status, and genetic information.. A detailed legal analysis is beyond the scope of this report.<sup>102</sup>

The Obama Administration big data report recommended using existing agencies and existing laws to combat discriminatory activities involving big data analytics, although it is not so clear that successor administrations would undertake the same type of responses.

**RECOMMENDATION:** The federal government’s lead civil rights and consumer Protection agencies, including the Department of Justice, the Federal Trade Commission, the Consumer Financial Protection Bureau, and the Equal Employment Opportunity Commission, should expand their technical expertise to be able to identify practices and outcomes facilitated by big data analytics that have a discriminatory impact on protected classes, and develop a plan for investigating and resolving violations of law in such cases. In assessing the potential concerns to address, the agencies may consider the classes of data, contexts of collection, and segments of the population that warrant particular attention, including for example genomic information or information about people with disabilities.<sup>103</sup>

The same report has other recommendations not immediately relevant here. For example, the report promoted the Obama Administration’s Consumer Bill of Rights, a legislative proposal that received little attention in earlier Congresses and that appear to be nothing more today than an historical footnote.<sup>104</sup> The report also discussed ways to make data resources more widely and more usefully available. Those are not the primary concern here, but better data sharing and better privacy protection can be compatible.

---

<sup>99</sup> Executive Office of the President, BIG DATA: SEIZING OPPORTUNITIES, PRESERVING VALUES at pages 58-59 (2014 Obama),

[https://obamawhitehouse.archives.gov/sites/default/files/docs/big\\_data\\_privacy\\_report\\_may\\_1\\_2014.pdf](https://obamawhitehouse.archives.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf).

<sup>100</sup> 42 U.S.C. § 2000e et seq., <https://www.law.cornell.edu/uscode/text/42/2000e>.

<sup>101</sup> Solon Barocas and Andrew D. Selbst, Big Data’s Disparate Impact, 104 Calif. L. Rev. 671 (2016), <http://www.californialawreview.org/wp-content/uploads/2016/06/2Barocas-Selbst.pdf>.

<sup>102</sup> See generally, Federal Trade Commission, BIG DATA A Tool for Inclusion or Exclusion? Understanding the Issues (2016), <https://www.ftc.gov/system/files/documents/reports/big-data-tool-inclusion-or-exclusion-understanding-issues/160106big-data-rpt.pdf>.

<sup>103</sup> Executive Office of the President, BIG DATA: SEIZING OPPORTUNITIES, PRESERVING VALUES at page 65 (2014 Obama),

[https://obamawhitehouse.archives.gov/sites/default/files/docs/big\\_data\\_privacy\\_report\\_may\\_1\\_2014.pdf](https://obamawhitehouse.archives.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf).

<sup>104</sup> Id. at page 61.

The PCAST report that emerged at the same time and from the same White House offered some other thoughts. It discussed at a high level of generality ways to regulate commerce using data analytics unless the activities are consistent with privacy preferences and community values. The discussion is interesting, but it received scant congressional or other attention.

Big data's "products of analysis" are created by computer programs that bring together algorithms and data so as to produce something of value. It might be feasible to recognize such programs, or their products, in a legal sense and to regulate their commerce. For example, they might not be allowed to be used in commerce (sold, leased, licensed, and so on) unless they are consistent with individuals' privacy elections or other expressions of community values (see Sections 4.3 and 4.5.1). Requirements might be imposed on conformity to appropriate standards of provenance, auditability, accuracy, and so on, in the data they use and produce; or that they meaningfully identify who (licensor vs. licensee) is responsible for correcting errors and liable for various types of harm or adverse consequence caused by the product.<sup>105</sup>

Major impediments to any discussion of responses include: 1) a lack of agreement on what is big data and big data analytics; 2) a lack of facts about existing industry or government activities that use big data analytics; 3) a lack of tools to measure discriminatory or other unwelcome consequences; and 4) an absence of generally applicable statutory privacy standards to apply in a new environment.

Taking a slightly different tack, it is useful to consider the recent Report of the Commission on Evidence-Based Policymaking.<sup>106</sup> The Commission did not address big data per se, but its charge was to explore the efficient creation of rigorous evidence as a routine part of government operations and the use of that evidence to construct effective public policy. The report offered recommendations for improving secure, private, and confidential data access. These recommendations overlap in some ways with the objectives for balanced use of big data, and some of the technical discussions and ideas are considered elsewhere in this report. In general, however, the Commission focused on government data, and much PII processed by federal agencies is subject to a privacy law. For unregulated health data, it is the private sector that is the focus of much concern. If implemented, however, some ideas of the Evidence-Based Policymaking Commission may have broader value.

A previous report from HHS covers the same ground as this discussion of big data, including attention to issues of unregulated health data. The Health IT Policy Committee's (HITPC) Privacy and Security Workgroup issued a report in August 2015 titled HEALTH BIG DATA RECOMMENDATIONS.<sup>107</sup> The recommendations generally point out problems, call for policymakers and others to do better, and support education and voluntary codes of conduct.

---

<sup>105</sup> President's Council of Advisors on Science and Technology, Big Data and Privacy: A Technological Perspective at page 49 (Obama 2014), [https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast\\_big\\_data\\_and\\_privacy\\_-\\_may\\_2014.pdf](https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_big_data_and_privacy_-_may_2014.pdf).

<sup>106</sup> Commission of Evidence-Based Policymaking, The Promise of Evidence-Based Policymaking (2017), <https://cep.gov/content/dam/cep/news/2017-09-06-news.pdf>.

<sup>107</sup> [https://www.healthit.gov/sites/faca/files/HITPC\\_Health\\_Big\\_Data\\_Report\\_FINAL.pdf](https://www.healthit.gov/sites/faca/files/HITPC_Health_Big_Data_Report_FINAL.pdf).

The report’s summary of recommendations follows.

### **6.1 Address Harm, Including Discrimination Concerns**

- ONC and other federal stakeholders should promote a better understanding by the public of the full scope of the problem – both harm to individuals and communities.
- Policymakers should continue to focus on identifying gaps in legal protections against what are likely to be an evolving set of harms from big data analytics.
- Policymakers should adopt measures that increase transparency about actual health information uses.
- Policymakers should explore ways to increase transparency around use of the algorithms used in big health analytics, perhaps with an approach similar to that used in the Fair Credit Reporting Act (FCRA).

### **6.2 Address Uneven Policy Environment**

- Promote Fair Information Practice Principles (FIPPs)-based protections for data outside of HIPAA:
  - o Voluntarily adopt self-governance codes of conduct. In order to credibly meet the requirements of both protecting sensitive personal information and enabling its appropriate use. Codes must include transparency, individual access, accountability, and use limitations.
  - o U.S. Department of Health and Human Services (HHS), Federal Trade Commission (FTC), and other relevant federal agencies should guide such efforts to more quickly establish dependable “rules of the road” and to ensure their enforceability in order to build trust in the use of health big data.
- Policymakers should evaluate existing laws, regulations, and policies (rules) governing uses of data that contribute to a learning health system to ensure that those rules promote responsible re-use of data to contribute to generalizable knowledge.
- Policymakers should modify rules around research uses of data to incentivize entities to use more privacy-protecting architectures, for example by providing safe harbors for certain behaviors and levels of security.
  - To support individuals’ rights to access their health information, create a “right of access” in entities not covered by HIPAA as part of the voluntary codes of conduct; also revise HIPAA over time to enable it to be effective at protecting health data in the digital age.
  - Educate consumers, healthcare providers, technology vendors, and other stakeholders about the limits of current legal protection; reinforce previous PSWG recommendations.
    - o Leverage most recent PSWG recommendations on better educating consumers about privacy and security laws and uses of personal information both within and outside of the HIPAA environment.

### **6.3 Protect Health Information by Improving Trust in De-Identification Methodologies and Reducing the Risk of Re-Identification**

- The Office for Civil Rights (OCR) should be a more active “steward” of HIPAA deidentification standards.
  - o Conduct ongoing review of methodologies to determine robustness and recommend updates to methodologies and policies.
  - o Seek assistance from third-party experts, such as the National Institute of Standards and Technology (NIST).
- Programs should be developed to objectively evaluate statistical methodologies to vet their capacity for reducing risk of re-identification to “very low” in particular contexts.

### **6.4 Support Secure Use of Data for Learning**

- Develop voluntary codes of conduct that also address robust security provisions.
- Policymakers should provide incentives for entities to use privacy-enhancing technologies and privacy-protecting technical architectures.
- Public and private sector organizations should educate stakeholders about cybersecurity risks and recommended precautions.
- Leverage recommendations made by the Privacy and Security Tiger Team and endorsed by the HITPC in 2011 with respect to the HIPAA Security Rule.
- Public and private sector organizations should educate stakeholders about cybersecurity risks and recommended precautions.
- Leverage recommendations made by the Privacy and Security Tiger Team and endorsed by the HITPC in 2011 with respect to the HIPAA Security Rule.

Other sources of ideas for responses to privacy concerns raised by big data may come from activities in artificial intelligence (AI). Like big data, AI lacks a clear consensus definition.<sup>108</sup> However, big data seems destined to be input to AI.<sup>109</sup> AI expansion is due to better algorithms, increases in networked computing power, and the ability to capture and store massive amounts of data. A recent report from the AI Now Institute at New York University is noteworthy.<sup>110</sup> Many of the topics raised in that report involve substantive applications of AI and the problems that arise.

While privacy is not the focus of the AI report, the report observes that privacy rights represent a “a particular sensitive challenge” for AI applications, especially in health care. Vulnerable populations may face increase risks.<sup>111</sup> The report includes a series of recommendations for government agencies, private companies, and others to address issues relating to lack of

---

<sup>108</sup> See What is Artificial Intelligence? An Informed Definition (2016), techemergence, (“One of the reasons AI is so difficult to define is because we still don’t have a set definition or one solid concept for intelligence in general.”), <https://www.techemergence.com/what-is-artificial-intelligence-an-informed-definition/>.

<sup>109</sup> See Anas Baig, Merging big data and AI is the next step (2017), TheNextWeb, <https://thenextweb.com/contributors/2017/08/19/merging-big-data-ai-next-step/>.

<sup>110</sup> AI Now 2017 Report, [https://assets.contentful.com/8wprhhvnpfc0/1A9c3ZTCZa2KEYM64Wsc2a/8636557c5fb14f2b74b2be64c3ce0c78/AI Now Institute 2017 Report .pdf](https://assets.contentful.com/8wprhhvnpfc0/1A9c3ZTCZa2KEYM64Wsc2a/8636557c5fb14f2b74b2be64c3ce0c78/AI%20Now%20Institute%202017%20Report.pdf).

<sup>111</sup> Id. At page 4.

transparency, potential biases from AI applications, more standards, more research, and the need for the development and application of ethical codes. Similar ideas may have some utility for the world of unregulated health information.

### III. Personal devices and the Internet of Things

#### A. Introduction

Personal devices and Internet of Things (IoT) devices bring to the table old issues (what is health information?) and new players from outside the health care world (device manufacturers, Internet services, consumer data companies, and others). The number of potential devices (personal or IoT) is enormous and increasing. Personal devices that collect health information include thermometers, pulse oximeters, blood pressure cuffs, clothing, belts, shoes, glasses, watches, activity monitors, cell phones, and many more. Almost any type of appliance, fitness equipment, camera, or other consumer product can become an IoT device with the capability of recording and reporting personal information over the Internet. An IoT device can collect data about activities, weight, health status, food purchases, eating habits, sleeping patterns, sexual activity, reading and viewing habits, and more. Some consumers undertake extensive self-reporting of their activities using devices and app of all types.<sup>112</sup> Considered as a whole, devices can collect a nearly unlimited assortment of data about individuals, and some of that data will be health information of some type. Some personal devices produce patient-generated health data (PGHD) of interest to the health care system, and some will not. Some devices will produce both PGHD and other data at the same time. For some data, determining its relevance to health care may be hard to tell, and today's answer may change tomorrow. Sorting out the categories returns to the definitional issue for health information.

It is unquestionable that many consumers embrace the use of these devices and that there are significant benefits from their use. A recent report from a public interest group, while acknowledging the promise of mobile and wearable devices, sees health information from those devices entering the growing digital health and marketing systems with the prospect of monetizing the data and the possibility of harms to consumers.

But some of the very features that make mobile and wearable devices so promising also raise serious concerns. Because of their capacity to collect and use large amounts of personal data—and, in particular, sensitive health data—this new generation of digital tools brings with it a host of privacy, security, and other risks. Many of these devices are already being integrated into a growing digital health and marketing ecosystem, which focuses on gathering and monetizing personal and health data in order to influence consumer behavior. As the use of trackers, smart watches, Internet-connected clothing, and other wearables becomes more widespread, and as their functionalities become even more sophisticated, the extent and nature of data collection will be unprecedented. Biosensors will routinely be able to capture not only an individual's heart rate, body temperature, and movement, but also brain activity, moods, and emotions. These data can, in turn, be combined with personal information from other sources—including health-care

---

<sup>112</sup> For more about lifelogging or the quantified self, see, e.g., *The Quantified Self: Counting Every Moment*, The Economist (2012), <http://www.economist.com/node/21548493/>. For a study of a particular type of health tracking, see, e.g., Daniel A. Epstein et al., *Examining Menstrual Tracking to Inform the Design of Personal Informatics Tools*, Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems pages 6876-6888 (2017), [http://www.depstein.net/pubs/depstein\\_chi17.pdf](http://www.depstein.net/pubs/depstein_chi17.pdf).

providers and drug companies—raising such potential harms as discriminatory profiling, manipulative marketing, and data breaches.<sup>113</sup>

The distinction between HIPAA regulated PHI and unregulated health data helps in sorting some things out. However, the permutations of players and regulations are many, complicated, and multidimensional. For example, Anna Slomovic points out that the integration of devices and apps in wellness programs can result in data flows to device and app makers, analytics companies, and in some cases social networks and marketers.<sup>114</sup> The same may be true for any unregulated health information originating from a device. Once consumer information enters the commercial data ecosystem, the information can end up almost anywhere and often be used for any purpose without time or other limit.

Regardless of the source, any identifiable consumer data that reaches a HIPAA covered entity is PHI and subject to regulation as other HIPAA PHI. That data is outside the scope of this report. If a covered entity sponsors a data collection activity (e.g., by giving a patient a device that reports to the covered entity), the device provider is likely to be a business associate of the covered entity and subject to HIPAA as well. The data remains PHI from source to file in the hands of all covered entities. If the patient also has access to the data (e.g., from the device), the data is not PHI in the hands of the patient. Patients can, of course, use their own health data as they see fit, whether the data is PHI elsewhere or not. Their options include sharing the data with health care providers or with unregulated third parties.

When the device manufacturer or device supporter is not a HIPAA covered entity or a business associate, HIPAA requirements do not attach. Any data produced by the patient and the device is unregulated health information in the hands of a manufacturer or any intermediary. The only applicable privacy protections are likely to derive from a privacy policy (if any) adopted by the device manufacturer, a policy likely to be subject to change by that manufacturer at any time.<sup>115</sup>

The Internet of Things (IoT) refers to the connection of almost any type of device (thermostat, toaster, refrigerator, etc.) can connect to the Internet, where the device can communicate with other devices, systems, or networks. Medical devices, such as infusion pumps, were once standalone instruments that interacted only with the patient or health care provider. These devices now can connect wirelessly to a variety of systems, networks, and other tools within a healthcare delivery organization (HDO) – ultimately contributing to the Internet of Medical Things (IoMT).<sup>116</sup>

---

<sup>113</sup> Center for Digital Democracy, *Health Wearable Devices in the Big Data Era: Ensuring Privacy, Security, and Consumer Protection* (undated, released in 2017), [https://www.democraticmedia.org/sites/default/files/field/public/2017/aucdd\\_wearablesreport\\_final121516.pdf](https://www.democraticmedia.org/sites/default/files/field/public/2017/aucdd_wearablesreport_final121516.pdf).

<sup>114</sup> Anna Slomovic, *eHealth and Privacy in U.S. Employer Wellness Programs*, in Ronald Leenes, Nadezhda Purtova, Samantha Adams (eds.), *Under Observation - The Interplay between eHealth and Surveillance* at page 13 (2017), [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2613452](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2613452).

<sup>115</sup> It is common for data controllers who have a privacy policy to reserve the right to change the policy at any time and often without any advance notice. Indeed, HIPAA *requires* that a privacy notice include a statement that the covered entity reserves the right to change the terms of the notice and to make the new provisions effective for all PHI, even PHI previously collected. 45 C.F.R. § 164.520(b)(1)(v)(C).

<sup>116</sup> See testimony of Kevin Stine, Chief Applied Cybersecurity Division, National Institute of Standards and Technology's Information Technology Laboratory, at the virtual hearing of the NCVHS Privacy, Confidentiality, and Security Subcommittee, (Nov. 28, 2017), ("The goal of our effort related to securing wireless infusion pumps at

Many industries, including healthcare, use or will use IoT facilities to collect data and health data. It seems unquestioned that many IoT devices bring useful capabilities. It also seems unquestioned that IoT allows for intrusive spying on individuals and a new class of security problems. In this regard, the IoT is no different than most information technologies. Personal devices, medical devices, and IoT devices may be somewhat undistinguishable from a privacy policy perspective even if the privacy regulatory rules differ.

There are further complexities here. Imagine a device provided by a covered entity that reports PHI in real time over the Internet to a physician. Properly encrypted data should be protected against intermediaries. If, however, a device does not encrypt data, then the patient's broadband provider could read data from the device as it passes over from the device, through the patient's router, and over to the provider's network. Intercepted data would not be PHI in the hands of the broadband provider.<sup>117</sup> Hackers can easily intercept some data from IoT devices that transmit data over the Internet without encryption.<sup>118</sup> Device data, like other Internet data, is subject to interception in different ways.

## B. Some Sources of Rules and Standards

### 1. Food and Drug Administration

The FDA regulates many but not all medical devices. In the mobile medical app space, for example, the FDA regulates some apps as a medical device. However, even though other mobile apps may meet the definition of a medical device, the FDA exercises enforcement discretion not to regulate them because they pose a lower risk to the public.<sup>119</sup> This leaves three categories of mobile apps that process health information: regulated medical devices, unregulated medical devices, and devices that are not medical devices at all. The resulting patchwork<sup>120</sup> of regulation echoes the way in which HIPAA regulates some health data while other health data remains

---

the center is to help healthcare providers secure these medical devices. Not just the devices, themselves, but really taking a view of the enterprise network or the enterprise system, if you will, where these devices reside, again, with a particular focus on the wireless infusion pumps. “), <https://www.ncvhs.hhs.gov/transcripts-minutes/transcript-of-the-november-28-2017-meeting-of-the-privacy-confidentiality-and-security-subcommittee/>.

<sup>117</sup> The Federal Communications Commission adopted a broadband privacy rule that would have required broadband providers to obtain opt-in consent to use or share health information. <https://www.fcc.gov/document/fcc-adopts-broadband-consumer-privacy-rules>. The rule did not include a definition of *health* information. Congress effectively repealed the rule before it took effect. Public Law No. 115-22 (2017), <https://www.gpo.gov/fdsys/pkg/PLAW-115publ22/html/PLAW-115publ22.htm>.

<sup>118</sup> See, e.g., Associated Press, Some top baby monitors vulnerable to hackers (2015), at CBS News.com, <https://www.cbsnews.com/news/baby-monitors-connect-internet-vulnerable-hackers-cybersecurity/>.

<sup>119</sup> Food and Drug Administration, Mobile Medical Applications Guidance for Industry and Food and Drug Administration Staff at page 23 (2015),

<http://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM263366.pdf>.

<sup>120</sup> Y. Tony Yang and Ross D. Silverman, Mobile Health Applications: The Patchwork Of Legal And Liability Issues Suggests Strategies To Improve Oversight, 33 Health Affairs (2014): 222–227 (2014), <http://content.healthaffairs.org/content/33/2/222.abstract>.

unregulated. The FDA has a similar approach to medical devices, regulating some and not others.<sup>121</sup>

A device regulated by the Food and Drug Administration is subject to FDA rules and guidance on cybersecurity.<sup>122</sup> If we presume proper cybersecurity, interception of the data is not a concern for a FDA regulated device. Unregulated medical devices and non-medical devices have no applicable rules on cybersecurity. For privacy, however, FDA does not require privacy safeguards and does not have rules or policies establishing standards for collection, use, and disclosure of any class of health information.

## 2. NIST

Researchers at the National Institute for Standards and Technology (NIST), part of the U.S. Department of Commerce, facilitate the development and adoption of standards for medical device communications for healthcare.<sup>123</sup> The standards support interoperability, security, and more. That is just one example of NIST standards relevant to the processing of health information. A HIPAA security website maintained by HHS references ten NIST publications on security matters of interest to HIPAA covered entities.<sup>124</sup> An FAQ states that use of these NIST standards is not a requirement of the security rule.<sup>125</sup> FDA also cites NIST standards as useful (but nonbinding) guidance for medical device manufacturers.<sup>126</sup>

To the extent that NIST standards apply to or are employed in activities of HIPAA covered entities, the data collected, transmitted, and stored pursuant to the standards is PHI. That PHI remains subject to the HIPAA security rule while in the hands of covered entities. It also falls outside the scope of this report. Yet activities of NIST like its ongoing work on the security of infusion pumps may be instructive for IoT devices not subject to HIPAA. At the November 28, 2017 virtual hearing of the NCVHS Privacy, Confidentiality, and Security Subcommittee, Kevin Stine, Chief Applied Cybersecurity Division, NIST's Information Technology Laboratory

---

<sup>121</sup> See, e.g., Food and Drug Administration, Medical Devices: Class I/II Exemptions, <https://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/Overview/ClassifyYourDevice/ucm051549.htm>.

<sup>122</sup> See, e.g., Food and Drug Administration, Content of Premarket Submissions for Management of Cybersecurity in Medical Devices Guidance for Industry and Food and Drug Administration Staff (2014), <https://www.fda.gov/downloads/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm356190.pdf>. See also <https://www.fda.gov/MedicalDevices/DigitalHealth/ucm373213.htm>.

<sup>123</sup> See National Institute of Standards and Technology, Medical Devices, <https://www.nist.gov/itl/ssd/medical-devices>.

<sup>124</sup> HHS.gov, Security Rule Guidance Material, <https://www.hhs.gov/hipaa/for-professionals/security/guidance/index.html>.

<sup>125</sup> HHS.gov, Are covered entities required to use the National Institute of Standards and Technology (NIST) guidance documents referred to in the preamble to the final Security Rule (68 Fed. Reg. 8334 (February 20, 2003))?, <https://www.hhs.gov/hipaa/for-professionals/faq/2015/are-covered-entities-required-to-use-the-nist-guidance-documents/index.html>.

<sup>126</sup> See, e.g., Food and Drug Administration, News Release, FDA outlines cybersecurity recommendations for medical device manufacturers (Jan. 15, 2016), <https://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm481968.htm>.

discussed how NIST's resources, standard guides, and practices are frequently voluntarily adopted by non-federal organizations.<sup>127</sup>

Of course, there is no reason why the same technology cannot accomplish similar purposes for unregulated health data. HIPAA Rules have the ability to mandate compliance by covered entities with NIST or other technical standards, but there is no current mechanism that requires others who process unregulated health data through devices or otherwise to conform to security standards or to technical standards.

### 3. Federal Trade Commission

The FTC has broad jurisdiction under its general power to take action against unfair and deceptive trade practices. Many commercial entities engaged in processing health information that is not PHI fall under the FTC's jurisdiction. However, the FTC has no practical ability to write rules except where Congress expressly directs the Commission to act. Two Commission actions are most relevant here.<sup>128</sup>

First, in 2009, Congress directed the Commission to issue a health breach notification rule for non-HIPAA providers of personal health records (PHR).<sup>129</sup> At the same time, HHS issued a similar rule for PHRs subject to HIPAA.<sup>130</sup> The Commission's rule applies to foreign and domestic vendors of PHRs, PHR related entities, and third party service providers, irrespective of any jurisdictional tests in the FTC Act.<sup>131</sup> This law and rule illustrate how the Commission can regulate non-HIPAA health information providers and can go beyond the limits in statute that restrict FTC jurisdiction and rulemaking authority. In the years since 2009, and despite the growth in health devices and apps, Congress has shown no interest in directing the Commission to issue additional rules in the health information space.

Second, the FTC produced a tool aimed at for mobile app developers to help determine how federal law applies to their products.<sup>132</sup> FTC developed the tool in cooperation with HHS, ONC, OCR, and FDA. The laws covered are HIPAA, the Federal Food, Drug, and Cosmetic Act, the Federal Trade Commission Act, and the FTC's Health Breach Notification Rule. The tool includes a link to a document containing the FTC's Best Practices for Mobile App Developers on

<sup>127</sup> See generally the testimony of Kevin Stine, Chief Applied Cybersecurity Division, National Institute of Standards and Technology's Information Technology Laboratory, at the virtual hearing of the NCVHS Privacy, Confidentiality, and Security Subcommittee, (Nov. 28, 2017), <https://www.ncvhs.hhs.gov/transcripts-minutes/transcript-of-the-november-28-2017-meeting-of-the-privacy-confidentiality-and-security-subcommittee/>.

<sup>128</sup> Other FTC responses include a sponsored contest for IoT security. See Federal Trade Commission Press Release, FTC Announces Winner of its Internet of Things Home Device Security Contest (July 26, 2017), <https://www.ftc.gov/news-events/press-releases/2017/07/ftc-announces-winner-its-internet-things-home-device-security>.

<sup>129</sup> Section 13047 of the American Recovery and Reinvestment Act of 2009, Public Law No. 111-5 (2009), [http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=111\\_cong\\_bills&docid=f:h1enr.pdf](http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=111_cong_bills&docid=f:h1enr.pdf). The law is also known as the Health Information Technology for Economic and Clinical Health Act (HITECH Act).

<sup>130</sup> 45 C.F.R. Part 164, Subpart D.

<sup>131</sup> Federal Trade Commission, Health Breach Notification Rule, 16 C.F.R. Part 318, <https://www.ecfr.gov/cgi-bin/text-idx?SID=9aed15bb4d4a308418d2e71eb1f5e1da&mc=true&node=pt16.1.318>.

<sup>132</sup> Federal Trade Commission, Mobile Health Apps Interactive Tool, <https://www.ftc.gov/tips-advice/business-center/guidance/mobile-health-apps-interactive-tool>.

privacy and security.<sup>133</sup> The Best Practices document poses a short series of simple questions suggesting issues for mobile app developers to consider when developing app privacy and security policies. There are currently no privacy rules specific to health applications outside of the scope of HIPAA and the FTC Act.<sup>134</sup>

#### 4. Industry and Other Standards

Industry standards and self-regulation are a recognized and sometimes controversial<sup>135</sup> way of developing privacy and security standards. Both of these activities typically rely on industry representatives, with little or no meaningful participation by consumer or privacy advocates. Still, industry activities on privacy standards and best practices can be a step in the right direction by calling attention to the need to consider privacy in developing products.

The Future of Privacy Forum (FPF) describes itself as a “nonprofit organization that serves as a catalyst for privacy leadership and scholarship, advancing principled data practices in support of emerging technologies.”<sup>136</sup> FPF proposed best practices for several technologies that raise privacy issues.<sup>137</sup> A 2016 document produced by FPF covers Best Practices for Consumer Wearables & Wellness Apps & Devices.<sup>138</sup>

Assessing the FPF standards is beyond the scope of this report, but the FPF document does address the definitional issue for health information. In an introductory section, the FPF document discusses the difficulty of drawing clear lines between health and lifestyle data, suggesting that treating all health-related personal data the same would be a mistake.

Given the lack of bright lines between sensitive health and non-sensitive lifestyle data, treating all health-related personal data the same would be a mistake. The stringent privacy, security, and safety requirements appropriate for medical devices and medical data would render many commercial fitness devices impractical for everyday consumers. At the same time, it would be a mistake to treat wellness data as if it were generic personal information without any sensitivity.<sup>139</sup>

---

<sup>133</sup> <https://www.ftc.gov/tips-advice/business-center/guidance/mobile-health-app-developers-ftc-best-practices>.

<sup>134</sup> At the November 28, 2017 virtual hearing of the NCVHS Privacy, Confidentiality, and Security Subcommittee, Frank Pasquale, Professor of Law, University of Maryland, observed that the 2009 HITECH Act placed a lot of responsibility on the FTC to look after unfairness or deceptiveness for transfers of data outside the HIPAA-protected zone and that the FTC does not have the staff “to keep up with it all.” <https://www.ncvhs.hhs.gov/transcripts-minutes/transcript-of-the-november-28-2017-meeting-of-the-privacy-confidentiality-and-security-subcommittee/>.

[The uncorrected transcript quotes Professor Pasquale as discussing the FCC, but he spoke about the FTC.]

<sup>135</sup> See, e.g., World Privacy Forum, *Many Failures: A Brief History of Privacy Self Regulation* (2011), <http://www.worldprivacyforum.org/wp-content/uploads/2011/10/WPFselfregulationhistory.pdf>.

<sup>136</sup> <https://fpf.org/about/>.

<sup>137</sup> <https://fpf.org/best-practices/>. FPF also maintains a useful index of best practices developed by other organizations, including government agencies, public interest groups, trade associations, and others.

<sup>138</sup> <https://fpf.org/wp-content/uploads/2016/08/FPF-Best-Practices-for-Wearables-and-Wellness-Apps-and-Devices-Final.pdf>.

<sup>139</sup> *Id.* at 2.

The FPF document is relevant, but its discussion and standards may lean on the side of industry supporters of FPF, and it is unclear how much weight it deserves. Still, FPF deserves credit for confronting hard issues.

The Consumer Technology Association (formerly the Consumer Electronic Association) is a technology trade association representing the U.S. consumer electronics industry. Its activities include developing technical standards for interoperability of devices and other technical matters. In 2015, it published “Guiding Principles on the Privacy and Security of Personal Wellness Data.”<sup>140</sup> The document is shorter and less detailed than the FPF

The Center for Democracy and Technology (CDT), a public interest group, teamed with Fitbit, a manufacturer of activity trackers, to produce a report offering practical guidance on privacy-protective and ethical internal research procedures at wearable technology companies.<sup>141</sup> The joint report did not propose standards for the day-to-day operations of a fitness tracker company but focused instead on developing “guidelines to preserve the dignity both of employees when they offer their personal data for experiments and for users whose data is involved throughout the R&D process.”<sup>142</sup>

Another potential “standard” that receives more attention today is *Privacy by Design*. The idea behind Privacy by Design is that developers should think about privacy from the beginning and minimize data collection.<sup>143</sup> The notion is sometimes associated with *privacy by default*. While those are welcome ideas generally, Privacy by Design/default typically has little substance or process associated with it other than a general exhortation to pay early attention to traditional approaches to minimize privacy consequences. The broad goals overlap in many ways with fair information practices, but mostly lack their specificity.

A privacy impact assessments (PIA), sometimes called a data protection impact assessment is a process rather than a standard. A PIA is a methodology for assessing the impact on privacy of a project, policy, program, service, product, or other initiative that involves the processing of personal information and, in consultation with stakeholders, for taking remedial action to avoid or minimize negative impacts.<sup>144</sup> The E-Government Act of 2002 requires PIAs of all federal agencies that develop or procure new information technology involving the collection, maintenance, or dissemination of information in identifiable form or that make substantial changes to existing information technology that manages information in identifiable form.<sup>145</sup> There is a similar requirement in the EU’s General Data Protection Regulation applicable to both public and private sector data controllers.<sup>146</sup>

<sup>140</sup> <http://www.ce.org/healthprivacy>.

<sup>141</sup> Center for Democracy and Technology and Fitbit, Inc., *Toward Privacy Aware Research and Development in Wearable Health* (2016), <https://cdt.org/insight/cdt-fitbit-report-privacy-practices-rd-wearables-industry/>.

<sup>142</sup> *Id.* at page 14.

<sup>143</sup> See, e.g., European Union, General Data Protection Regulation [2016] OJ L119/1, at Article 25, <http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52012PC0011>.

<sup>144</sup> See David Wright & Paul De Hert, Introduction to Privacy Impact Assessment at 5, in David Wright & Paul De Hert, editors, *Privacy Impact Assessment* (2012), <http://www.springer.com/us/book/9789400725423>.

<sup>145</sup> 44 U.S.C. § 3501 note, <https://www.law.cornell.edu/uscode/text/44/3501>.

<sup>146</sup> European Union, General Data Protection Regulation [2016] OJ L119/1, at Article 55, <http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52012PC0011>.

The content of and process for PIAs are variable around the world. The federal agency PIA requirement adds little to the existing requirement under the Privacy Act of 1974 for publication of a system of records notice for most agency collections of personal information.<sup>147</sup> Other types of PIAs in use around the world differ considerably in content, timing, and effectiveness, with little consensus about the best approach.<sup>148</sup> However, PIAs typically have substantive and procedural requirements, something less often found in connection with privacy by design. The notion of a PIA seems well-entrenched today as a privacy process. The jury is still out for Privacy by Design.

### C. Devices in Context

Many different types of devices and device-based functions are in use and it is not possible here to describe the full range of activities. Outside of devices that produce HIPAA PHI, many devices acquired and used by individuals create data that, in addition to any personal uses, likely enter the commercial marketplace for consumer data. This marketplace is largely unregulated for privacy. This section highlights a few device-based activities that illustrate interesting and complex applications in different spheres.

#### 1. Wellness Programs

Wellness programs may use a fitness device and offer a useful example of an activity that has the potential to develop non-regulated pools of health data.<sup>149</sup> Some wellness programs create HIPAA PHI. If an employer sponsors a wellness program through a health plan, the program (and the data) will be subject to the HIPAA rules because the plan is a covered entity and the program is a business associate. Participants in a wellness program (or other activity) subject to HIPAA can consent to the sharing of PHI with non-HIPAA entities, and if they do, their information “escapes” from HIPAA rules in the hands of the recipient. Employer access to and use of wellness program data

A wellness program (personal, employer, community, or otherwise) can easily be structured to avoid HIPAA so that no one who obtains health data has HIPAA obligations. Data so collected may be completely free from regulation. Of course, not all wellness program data comes from devices, but mobile devices can be both a data collection mechanism as well as a way to offer analytic results, feedback and recommendations to participants.<sup>150</sup>

Wellness activities in a workplace environment are interesting and different because there are Equal Employment Opportunity Commission rules under both the Americans With Disabilities Act and the Genetic Information Nondiscrimination Act that impose some limits on employer

---

<sup>147</sup> 5 U.S.C. § 552a(e)(4).

<sup>148</sup> See generally, See David Wright & Paul De Hert, Introduction to Privacy Impact Assessment at 5, in David Wright & Paul De Hert, editors, *Privacy Impact Assessment* (2012), <http://www.springer.com/us/book/9789400725423>.

<sup>149</sup> See generally, Anna Slomovic, eHealth and Privacy in U.S. Employer Wellness Programs, in Ronald Leenes, Nadezhda Purtova, Samantha Adams (eds.), *Under Observation - The Interplay between eHealth and Surveillance* (2017), [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2613452](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2613452).

<sup>150</sup> Id. at page 7.

collection and use of health information.<sup>151</sup> Many other activities using devices collect wholly unregulated health information that may be used for a wide variety of commercial purposes.

It is unclear whether employees or others participating in wellness programs understand the nature of any applicable privacy protections or even know of all the entities that obtain their data. The same conclusion is likely for many devices that collect consumer information or health information. Of course, many other devices and websites collect, use, and share other non-health with little actual knowledge by consumers of the scope of the processing of their data.

## 2. Citizen Science

Citizen science is a form of open collaboration in which members of the public participate in scientific research to meet real world goals. Crowdsourcing is a process by which individuals or organizations solicit contributions from a large group of individuals or a group of trusted individuals or experts. These activities occasionally involve the collection of information that might be considered health information. Both citizen science and crowdsourcing are typically noncommercial activities.

An example comes from the Health eHeart study at the University of California San Francisco.<sup>152</sup> The Health eHeart Study is an online study with the goal of stopping the progression of heart disease. Participants fill out surveys, connect mobile devices, and use cell phone apps to provide personal information. Some participants will use special sensors or their smartphone to track record pulse, weight, sleep, activity, behavior, and more. The privacy policy for the study limits data use to research, restricts non-consensual disclosures, and relies on the protections of the HHS certificate of confidentiality program.<sup>153</sup> These are generally appropriate policies for a program of this type, and this study may not be representative of the attention to privacy paid by all citizen science activities.

---

<sup>151</sup> Equal Employment Opportunity Commission, Regulations Under the Americans With Disabilities Act, 29 C.F.R. Part 1630, <https://www.ecfr.gov/cgi-bin/text-idx?SID=4e2e48165f43b03b761c31fe95d0d7f5&mc=true&node=pt29.4.1630&rgn=div>; Genetic Information Nondiscrimination Act, 29 C.F.R. Part 1635, <https://www.ecfr.gov/cgi-bin/text-idx?SID=4e2e48165f43b03b761c31fe95d0d7f5&mc=true&node=pt29.4.1635&rgn=div5>.

<sup>152</sup> <https://www.health-eheartstudy.org/study>.

<sup>153</sup> <https://www.health-eheartstudy.org/privacypolicy>.

## IV. Laws in Other Domains

### A. U.S. Privacy Model vs. the EU Privacy Model

Most of the world generally follows the European Union approach to regulating personal information for privacy.<sup>154</sup> The EU approach establishes a set of data protection rules and procedures for personal data that applies broadly to nearly all record keepers.<sup>155</sup> Fair Information Practices form the basis for EU data protection policies. Within the general framework, rules vary in their application depending on circumstances. For example, the rules generally prohibit the processing of health data. However, processing is allowed if one of ten circumstances applies.<sup>156</sup> The circumstances include processing for health care treatment, public health, scientific research, and more. For some types of health care processing, other standards, procedures, or laws may provide an additional set of rules or procedures.

For present purposes, the most important aspect of the EU approach is that rules apply to nearly all record keepers. As health information passes from hand to hand, each data controller remains subject to EU data protection rules. While there may be difference in application of the rules in some instances, basic rules apply to all.

The U.S. approach is different, with some describing it as sectoral. Privacy rules apply to some types of personally identifiable information or to some types of data held by some classes of record keepers. Much PII is subject to no privacy law at all. As regulated PII passes from one record keeper to another, privacy rules rarely follow the records.<sup>157</sup> In some circumstances, a receiving record keeper may be independently subject to the same set of privacy rules that apply to the record keeper disclosing the records. This is somewhat true for HIPAA, where records that pass from covered entity to covered entity are subject to the same HIPAA Rules because all covered entities must follow the HIPAA Rules.

However, when HIPAA records pass from a covered entity to a non-covered entity, either no privacy rules apply to the receiving record keeper or a different set of privacy rules apply. So if a HIPAA covered entity sends a health record to a school, the record in the hands of the school is subject to the Family Educational Rights and Privacy Act (FERPA) and not to HIPAA. If HIPAA records pass to a third party who is not a covered entity (e.g., a public health agency,

<sup>154</sup> See generally Graham Greenleaf, The influence of European data privacy standards outside Europe: Implications for Globalisation of Convention 108, <http://ssrn.com/abstract=1960299>.

<sup>155</sup> The EU promulgated its original data protection policy through a Data Protection Directive that required Member States to adopt national laws implementing the Directive's standards. Council Directive No. 95/46/EC of 24 October 1995 on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data, OJ L 281/31(1995), <http://eur-lex.europa.eu/legal-content/en/ALL/?uri=CELEX:31995L0046>. The General Data Protection Regulation (GDPR) that replaces the Directive takes effect in May 2018 establishes EU-wide law. Regulation (EU) 2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC, General Data Protection Regulation [2016] OJ L119/1, <http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52012PC0011>.

<sup>156</sup> European Union, General Data Protection Regulation [2016] OJ L119/1, at Article 9, <http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52012PC0011>.

<sup>157</sup> The principal example of a privacy rule that follows the record is 42 U.S.C. Part 2 covering the confidentiality of substance use disorder patient records, <https://www.ecfr.gov/cgi-bin/text-idx?SID=53f2e17e5ac033dfa412916c05c81f9e&mc=true&node=pt42.1.2&rgn=div5>.

researcher, law enforcement agency, or national security agency), HIPAA does not apply to the records in the hands of the recipient. If the recipient is a federal agency, the Privacy Act of 1974 will often but not always apply to the records. If the recipient is not a federal agency, no privacy law will apply.

What is important here is that a record containing health data is not always subject to a privacy rule. Much depends on who holds the records. The record may be subject to multiple privacy laws at the same time, or the record may be subject to no privacy protections at all.

For health records that originate with a record keeper other than a HIPAA covered entity, the records may never be subject to any privacy law in the hands of that record keeper. If the records pass from the originator to another record keeper, the records are in most instances free from privacy regulation unless they end up in the hands of a HIPAA covered entity. Thus, a record of the purchase of an over-the-counter drug typically falls under no privacy regulation in the hands of the seller of that drug. If the purchaser reports the purchase to a HIPAA covered entity, the information in the hands of that entity falls under the HIPAA privacy rule. The same information remains unregulated for privacy in the hands of the original seller.

Commercial companies that voluntarily adopt privacy policies can be held to compliance with those policies by the Federal Trade Commission. The FTC has authority to prevent companies that fall within its jurisdiction from engaging in unfair or deceptive trade practices.<sup>158</sup> Not complying with a published privacy policy can be a deceptive trade practice. The FTC has jurisdiction over many commercial entities, but its unfairness and deception authority does not extend to the insurance industry, banks, airlines, non-profits, and state and local governments.<sup>159</sup> Overall, FTC authority extends to roughly half of the economy.

As a practical matter, the FTC does not have the ability to issue general privacy regulations except in those areas where a statute expressly authorizes the Commission to act (e.g., the Children's Online Privacy Protection Act<sup>160</sup>).<sup>161</sup> The vagueness of the unfair and deceptive practices standard provides little specific direction to anyone seeking to implement privacy protection. FTC case law provides some general guidance, but no rules. The Commission can take stronger action against a company that signed a consent decree in a previous case, but the number of privacy and security cases it brings is relatively small (as compared to the number of cases brought by OCR at HHS). The Commission also issues a variety of advisory materials (e.g., staff reports) that provide guidance to industry.<sup>162</sup> The FTC has a large jurisdiction

---

<sup>158</sup> 15 U.S.C. § 45, <https://www.law.cornell.edu/uscode/text/15/45>.

<sup>159</sup> See Woodrow Hartzog and Daniel J. Solove, *The Scope and Potential of FTC Data Protection*, 83 *George Washington Law Review* 2230, 2289 (2015), [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2461096##](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2461096##). The scope of the Commission's authority over common carriers is in dispute at present, but the Commission traditionally did not exercise jurisdiction over common carriers. In recent years, the Consumer Finance Protection Bureau became active in some privacy matters.

<sup>160</sup> 15 U.S.C. § 6502(b), <https://www.law.cornell.edu/uscode/text/15/6502>.

<sup>161</sup> See Chris Jay Hoofnagle, *Federal Trade Commission Privacy Law and Policy* at 101 (2016).

<sup>162</sup> See, e.g., Federal Trade Commission, *Facing Facts: Best Practices For Common Uses of Facial Recognition Technologies* (2012) (Staff Report), <https://www.ftc.gov/reports/facing-facts-best-practices-common-uses-facial-recognition-technologies>. Despite the changes in facial recognition technology and use in recent years, the Commission did not update the guidance.

covering broad issues like consumer protection and antitrust and limited resources. The value of FTC activities in privacy is a controversial matter, and the Commission's interest in and commitment to privacy varies over time.

## **B. Fair Credit Reporting Act and Its Limits**

The Fair Credit Reporting Act (FCRA) principally regulates the collection, use, and dissemination of credit reports. The credit reporting industry collects information on consumer activities relevant to creditworthiness. Credit reports do not typically include health information, but they do include information on medical debt.<sup>163</sup> The FCRA restricts the use of credit reports with the main allowable activities (“principal purposes”) relating to employment, insurance, and credit.<sup>164</sup>

The FCRA is of interest here because the pressures of the information marketplace together with new technology undermine the goals of the Act. The law seeks to strike a balance between the legitimate interests of creditors, employers, and insurers in evaluating consumer credit and the need of consumers for fair treatment and transparency. If a creditor uses a credit report and takes an adverse action (e.g., denies a consumer credit), the consumer has rights under the FCRA. These rights include notice and the opportunity to challenge the credit report's accuracy. The FCRA also imposes obligations on those who furnish information about consumers to credit bureaus.

Today, other information about consumers analyzed and massaged by algorithms can produce the same types of results previously only attainable from regulated credit reports. At least one company claims that if your Facebook friends do not pay their bills on time, then it is likely that you won't pay your bills on time either.<sup>165</sup> This type of information does not come directly from credit reports, and its use may not create clear rights for consumers denied credit on the basis of information about their friends that was not collected for credit reporting purposes. How would a consumer fight a judgment made on the basis of information about others? The traditional remedies of the FCRA do not match up with the realities of the current marketplace.

The implications for the unregulated world of health data are similar. Like the FCRA, HIPAA gives patients a basket of rights. Both the FCRA and HIPAA implement Fair Information Practices. Under HIPAA, patients can see their health records and protect their own interests. HIPAA covered entities have obligation to consider patient requests for amendment. With the growing availability of non-regulated health data from disparate sources, merchant, marketers, profilers, and others can make increasingly sophisticated judgments about the health of consumers and act on those judgments without any obligation to give consumers any rights with respect to the processing of that data. In many ways, the lack of balance is not different from the

---

<sup>163</sup> As a result of a recent settlement with the New York Attorney General, the three major national credit bureaus agreed to adjust the way they report on medical debt, including a six month delay in reporting. See Attorney General of the State of New York, Bureau of Consumer Frauds & Protection, In the Matter of the Investigation by Eric T. Schneiderman, Attorney General of the State of New York, of Experian Information Solutions, Inc.; Equifax Information Services, LLC; and TransUnion LLC (2015), <https://ag.ny.gov/pdfs/CRA%20Agreement%20Fully%20Executed%203.8.15.pdf>.

<sup>164</sup> 15 U.S.C. § 1681b, <https://www.law.cornell.edu/uscode/text/15/1681b>.

<sup>165</sup> See Erika Eichelberger, Your Deadbeat Facebook Friends Could Cost You a Loan, Mother Jones (Sep. 18, 2013), <http://www.motherjones.com/politics/2013/09/lenders-vet-borrowers-social-media-facebook/>.

types of marketing activities that took place in the middle of the 20<sup>th</sup> century. Consumers then had little awareness of the mailing list industry and generally had no rights. Today, the amount of data available about individual consumers is much larger and more three dimensional than before. The availability of modeled data is also greater, as is the ability of algorithms to digest all that data and produce results.

One major difference between regulated and non-regulated health data is the interest in accuracy. Health care providers want data about patients that is as accurate as possible. Marketers, on the other hand, can still profit from inaccurate data. For example, a traditional measure for a good response to a commercial mailing is two percent.<sup>166</sup> If only 80% of the individuals sent a catalog actually meet all the criteria for selection, the mailing can still be profitable. If an employer knows that a candidate has a fifty percent chance of having an unwanted characteristic, the employer can simply hire someone else who has a lower chance of that characteristic.

There are more dimensions to the availability of unregulated health information. The Americans with Disabilities Act<sup>167</sup> (ADA) generally prohibits an employer from investigating an employee's medical condition beyond what is necessary to assess that individual's ability to perform the essential functions of a job or to determine the need for accommodation or time away from work. The ADA says in effect that an employer cannot ask if a prospective employee has diabetes. Looking at the records of a data profiler to determine if that individual is likely to be a diabetic would violate the principle that you cannot do indirectly what you are prohibited from doing directly.<sup>168</sup> Whether an employer who trolled Facebook to investigate job applicants would be "caught" is an open question.

Consider a data company that develops an algorithm that identifies with a reasonable degree of accuracy whether an individual is a diabetic. Using that algorithm to determine if a job applicant is a diabetic would violate the ADA. Now suppose that the algorithm finds several non-health characteristics (purchase of selected food items, use of specific over-the-counter medical products, interest in biofeedback, etc.) that correlate with the likelihood of diabetes. It is somewhat harder to conclude (and much harder to show) that an inquiry about those non-health characteristics violates the law.

A broader point is that the traditional legislative balancing efforts in the United States for the collection and use of personal information are losing touch with the challenges of the modern information technology age. This is a problem for personal data other than unregulated health data too. Solutions are hard to find, and debate about the shortcomings of the FCRA in today's information marketplace is occasional at best. Any proposal that might address non-regulated health data must initially confront the challenging problem of defining what constitutes "health" data.

---

<sup>166</sup> See Mcarthy & King Marketing, What Makes a Good Direct MailResponse Rate?, <http://www.mccarthyandking.com/direct-marketing-tutorials/learning-direct-mail-response-rates>.

<sup>167</sup> 42 U.S.C. § 12101 et seq., <https://www.law.cornell.edu/uscode/text/42/12101/>

<sup>168</sup> See Equal Employment Opportunity Commission, ADA Enforcement Guidance: Preemployment Disability-Related Questions and Medical Examinations, <https://www.eeoc.gov/policy/docs/preemp.html> ("May an employer ask third parties questions it could not ask the applicant directly? No. An employer may not ask a third party (such as a service that provides information about workers' compensation claims, a state agency, or an applicant's friends, family, or former employers) any questions that it could not directly ask the applicant.")

### C. Other Sources

Other countries wrestle with rules for making health data available while protecting privacy. A recent report from the Organisation for Economic Cooperation and Development (OECD) addresses health data governance, finding significant cross-country differences in data availability and use.<sup>169</sup> The OECD report finds that health data collected by national governments that can be linked and shared are a valuable resource that can be used safely to improve the health outcomes of patients and the quality and performance of health care systems. The report supports the development of privacy-protective uses of personal health data, identifying key data governance mechanisms that maximize benefits to patients and to societies and minimize risks to patient privacy and to public trust and confidence in health care providers and governments.

The report identifies eight key data governance mechanism:

1. The health information system supports the monitoring and improvement of health care quality and system performance, as well as research innovations for better health care and outcomes.
2. The processing and the secondary use of data for public health, research and statistical purposes are permitted, subject to safeguards specified in the legislative framework for data protection.
3. The public are consulted upon and informed about the collection and processing of personal health data.
4. A certification/accreditation process for the processing of health data for research and statistics is implemented.
5. The project approval process is fair and transparent and decision making is supported by an independent, multidisciplinary project review body.
6. Best practices in data de-identification are applied to protect patient data privacy.
7. Best practices in data security and management are applied to reduce re-identification and breach risks.
8. Governance mechanisms are periodically reviewed at an international level to maximise societal benefits and minimise societal risks as new data sources and new technologies are introduced.

Not all of these ideas are new to the US, and they likely reflect consensus goals here as well as in other countries. If there is a fly in this particular ointment for the US, it is that the world of unregulated health information seems beyond the reach of any of these governance mechanisms under existing law. Even if unregulated health data is useful for public health, research, and other beneficial purposes, that data stands outside the rules that guide these activities.

---

<sup>169</sup> Organisation for Economic Cooperation and Development, Health Data Governance PRIVACY, MONITORING AND RESEARCH (2015), <http://www.oecd.org/publications/health-data-governance-9789264244566-en.htm>. See also Organisation for Economic Cooperation and Development, Recommendation of the OECD Council on Health Data Governance (2017), <http://www.oecd.org/health/health-systems/Recommendation-of-OECD-Council-on-Health-Data-Governance-Booklet.pdf>.

At a September 13, 2017, NCVHC hearing, Fatemeh Khatibloo, a researcher and analyst Forrester Research, discussed the possibility of expanding the scope of HIPAA.

Now, this is only going to get more complicated in the future. Who has familiarity with the case of a gentleman's pacemaker actually being used to [indict] him for arson? Well, about six weeks ago, a judge decided that was admissible evidence in court. Here we have a situation where you may be sitting in a hospital bed talking to a cardiologist who says, you need a pacemaker to keep you alive. Except that data could be used against you in the future. What do you choose to do? It is not a fair question to ask of a patient.

So we think that there are six ways that we might think about HIPAA in the future. We think that the definition of protected health information should be expanded. It should be expanded to include wellness and behavior data. But it should be defined by class and the potential for harm and the sensitivity of that data. We think that we should require proportionately appropriate protection and handling of each class of data, not one broad set of data. And we should be limiting the use of sensitive data irrespective of the provider or practitioner. This isn't about these type of data or the covered entity. This is about the citizen's rights to have protected data.

We think that all firms that are collecting this PHI and health data should be subject to the same privacy and security standards. Most importantly, we think that we should be providing meaningful control over health care data to the individual, to the person about whom it is related, to whom it is related.<sup>170</sup>

It is clear from her testimony that she supports extending privacy protections to devices and apps that are currently beyond the scope of HIPAA. However, it is less clear how that goal might be accomplished. Aside from the administrative and political barriers that any expansion of HIPAA must overcome, there are policy and conceptual problems. HIPAA Privacy Rules work in the context of health care providers and insurers. The same rules may not work in the same way for other health data processors so that any simple extension of HIPAA may not be practical.

---

<sup>170</sup> <https://www.ncvhs.hhs.gov/transcripts-minutes/transcript-of-the-september-13-2017-ncvhs-full-committee-meeting/#hipaa>.

## V. Evolving technologies for privacy and security

For privacy and security, technology is a two-edged sword. Technology can both protect and erode privacy and security, with the same technology doing both things at the same time. An Internet filter can prevent a user from reaching websites loaded with malware, but at a cost of monitoring and recording all Internet activities. A device can track a user's health status, but it may report on the user's activities and location. A self-driving car may be a great enabler and convenience, but it may result in the recording of everywhere you go, when, and how often.

Dr. Jeremy Epstein, Deputy Division Director for Computer Network Systems at the National Science Foundation, offered an interesting example about how better methods of identity authentication may have unintended consequences for privacy.

So as an example, there is more going on to put a device on your phone or on your laptop that can authenticate who you are based on your unique heart rate. You don't have to identify yourself, or you don't have to provide a password or anything like that. Basically, your heart rate is enough. That is not what we typically think of as an authenticator. We typically think of passwords and fingerprints and stuff like that.

The more accurate that heart rate monitor is, and these are talking about using just an ordinary cell phone or similar device to be able to do that. The more accurate they are, the better they are from a security perspective, but the greater the risk from a privacy perspective. Similarly, people are doing things with brainwaves for authentication.<sup>171</sup>

Technological consequences, both intended and unintended, need to be assessed for their secondary effects on privacy. The Brandeis program at the Defense Advanced Research Projects Agency (DARPA) looks for technology that will support both privacy and the use data sharing at the same time.

The Brandeis program seeks to develop the technical means to protect the private and proprietary information of individuals and enterprises. The vision of the Brandeis program is to break the tension between: (a) maintaining privacy and (b) being able to tap into the huge value of data. Rather than having to balance between them, Brandeis aims to build a third option – enabling safe and predictable sharing of data in which privacy is preserved.<sup>172</sup>

Given the limits of this report, technologies within the health care system regulated under HIPAA Rules are not of primary interest. Yet, technology applications will not reflect or respect HIPAA boundaries. The same technology regulated for privacy and security in one context will lack any controls in another context. Given that the world of unregulated health data is, of

---

<sup>171</sup> National Committee on Vital and Health Statistics Full Committee (Sep. 13, 2017),

<https://www.ncvhs.hhs.gov/transcripts-minutes/transcript-of-the-september-13-2017-ncvhs-full-committee-meeting/>.

<sup>172</sup> <https://www.darpa.mil/program/brandeis>.

course, unregulated, it is difficult from a policy perspective to consider restraints, controls, or management of technology under current rules. One can, however, observe events and developments and hope to influence them or their adoption in some useful way.

Technology also makes some things harder. Nicole Gardner, Vice President of IBM's Services Group, testified at the September 13, 2017, hearing, about how IoT turns data that previously was mostly static into data that is in motion most of the time.

There are a lot of other questions about data because data is not actually a static thing. So when we talk about transactions, and we talk about transmitting information from one place to another in a kind of traditional way from one system to another over a third party telecommunications environment, we are talking about data that is generally at rest and in motion for just a little while.

But when you add in the Internet of Things, data all of a sudden becomes in motion most of the time. So there are no governance structures or policies or frameworks or agreements or legal boundaries or anything around data in motion. And the more that the volume increases, and the more we become living in a world of the Internet of Things, the more data in motion is going to become interesting and more of a challenge.<sup>173</sup>

It seems apparent that the privacy challenges of IoT and even more data in motion become that much greater.

The focus in this section is on current technologies in the context of unregulated health information activities and in conjunction with a discussion of policy problems that technology raises. This is a sample and by no means a complete review of current technologies or policies. A better source on statistical technologies is a recent report from the National Academy of Sciences which, while focused on federal statistics, reviews issues relating to privacy and technology to support better uses of data.<sup>174</sup> The report includes a recommendation that federal agencies should adopt privacy-preserving and privacy-enhancing technologies.<sup>175</sup> The report also concludes that better privacy protections can potentially enable the use of private-sector data for federal statistics.<sup>176</sup>

The notion that evolving technologies can protect privacy while making greater use of personal data may be just as applicable to the private sector and to unregulated health information in the hands of private-sector entities. However, a mechanism that would encourage or require use of those technologies by private-sector companies is not immediately apparent.

---

<sup>173</sup> National Committee on Vital and Health Statistics Full Committee (Sep. 13, 2017), <https://www.ncvhs.hhs.gov/transcripts-minutes/transcript-of-the-september-13-2017-ncvhs-full-committee-meeting/>.

<sup>174</sup> National Academy of Sciences, INNOVATIONS IN FEDERAL STATISTICS Combining Data Sources While Protecting Privacy (2017), <https://www.nap.edu/catalog/24652/innovations-in-federal-statistics-combining-data-sources-while-protecting-privacy>.

<sup>175</sup> Id. at 96 (recommendation 5-2).

<sup>176</sup> Id. at 64 (Conclusion 4-1).

## A. Applied technologies can get complicated quickly

The actual technology behind any particular device or methodology can be simple or complex in conception. A drone is a sensor in the sky that collects and reports pictures and data. That is a simple concept that can be understood by most without knowing any of the engineering details. A blockchain is a shared and distributed database with multiple identical copies, no central ownership, and management by a consensus of network participants who work together using cryptography to decide what can be added to the database.<sup>177</sup> Even if you clearly understand the basic idea of a blockchain, it is not so easy to understand how to employ blockchain technology in a health context.

Here is a basic and simple description of a blockchain application for medications. It is not that simple to follow the details, even at a high-level of abstraction.

One of the first use cases that typically pop up when discussing blockchain and health care is data exchange. Take medication prescribing as an example. A patient's medications are frequently prescribed and filled by different entities — hospitals, provider offices, pharmacies, etc. Each one maintains its own “source of truth” of medications for a patient, frequently with outdated or simply wrong information. As a result, providers in different networks, or on different EHRs, may not see one another's prescriptions. Additionally, electronic prescriptions must be directed to specific pharmacies, and paper prescriptions can be duplicated or lost.

To counter these difficulties, a medication prescription blockchain could be a shared source of truth. Every prescription event would be known and shared by those authorized to see it. This would allow, for example, prescriptions to be written electronically without specifying a pharmacy, or prescriptions to be partially filled (and “fully” filled at a later date, by a different pharmacy). Since the blockchain would be the source of truth, each pharmacy would see all events surrounding that prescription — and could act accordingly. Most importantly, all health care providers could have an immediate view into a patient's current medications, ensuring accuracy and fidelity.<sup>178</sup>

Nicole Gardner, Vice President of IBM's Services Group, testified at the September 13, 2017, hearing that IBM thought blockchain was so important that it formed an entire division around the use of Blockchain in the world of security and privacy.<sup>179</sup> There is clearly much potential for blockchain application.

---

<sup>177</sup> Blockchain has many potential applications in multiple areas. Blockchain is the technology that supports the digital currency Bitcoin.

<sup>178</sup> William Gordon, Adam Wright, & Adam Landman, *Blockchain in Health Care: Decoding the Hype* (Feb. 9, 2017), <https://catalyst.nejm.org/decoding-blockchain-technology-health/>. For a prototype of a blockchain application, see Ariel Ekblaw, Asaph Azaria, John D. Halamka, & Andrew Lippman, see *A Case Study for Blockchain in Healthcare: “MedRec” prototype for electronic health records and medical research data* (2016), [https://www.healthit.gov/sites/default/files/5-56-onc\\_blockchainchallenge\\_mitwhitepaper.pdf](https://www.healthit.gov/sites/default/files/5-56-onc_blockchainchallenge_mitwhitepaper.pdf).

<sup>179</sup> National Committee on Vital and Health Statistics Full Committee (Sep. 13, 2017), <https://www.ncvhs.hhs.gov/transcripts-minutes/transcript-of-the-september-13-2017-ncvhs-full-committee-meeting/>.

Implementing blockchain in the health care system with regulated health information in an environment with hundreds of thousands of health care providers, thousands of pharmacies, hundreds of oversight agencies, and hundreds of millions of patients would obviously be challenging, expensive, and take time. Implementing blockchain for a narrow part of the healthcare system (e.g., prescription drugs) will be easier than implementing a blockchain for the entire health care system. The basic notion of an electronic health record (EHR) is much simpler conceptually than a blockchain, but the implementation of EHRs has not been simple and has not achieved many of the intended objectives. Explaining blockchain to everyone who would need to understand it (including patients) would be difficult, to say the least. Much of the technological details can be hidden from patient view, but some of the promised benefits require patient involvement.

This is not to suggest that blockchain has no potential uses in health care.<sup>180</sup> Blockchain is the subject of attention from the regulated health care world. In 2016, the Office of the National Coordinator at HHS conducted a contest seeking papers suggesting new uses for Blockchain to protect and exchange electronic health information. The contest attracted more than 70 papers.<sup>181</sup> That activity takes place in the world of regulated health data.

At the NCVHS hearing, Dr. Jeremy Epstein, Deputy Division Director for Computer Network Systems at the National Science Foundation, offered an interesting observation and caution about the current enthusiasm for blockchain.

I wanted to comment about Blockchain for a second, if I may. When all you have is a hammer, everything looks like a nail, and I think that is where we are at with Blockchain today. People are using Blockchain to solve all the world's problems just because that is the hammer they have.<sup>182</sup>

In the world of unregulated health data, the challenge is greater. In the regulated health care system, patients, providers, and other participants know most of the players and likely have some sense of the flow of information. That knowledge is far from universal or perfect, of course. For example, many business associates are invisible to most patients.

When we consider unregulated health data, the number of categories of players is large and not completely known outside narrow industries. A 2007 NCVHS report identified health care entities not covered by HIPAA (cosmetic medicine services, occupational health clinics, fitness clubs, home testing laboratories, massage therapists, nutritional counselors, “alternative”

---

<sup>180</sup> See, e.g., Deloitte, Blockchain: Opportunities for Health Care (2016), [www.healthit.gov/sites/default/files/4-37-hhs\\_blockchain\\_challenge\\_deloitte\\_consulting\\_llp.pdf](http://www.healthit.gov/sites/default/files/4-37-hhs_blockchain_challenge_deloitte_consulting_llp.pdf).

<sup>181</sup> Press release, Office of the National Coordinator for Health Information Technology, ONC announces Blockchain challenge winners, Sep. 1, 2016, <http://wayback.archive-it.org/3926/20170127190114/https://www.hhs.gov/about/news/2016/08/29/onc-announces-blockchain-challenge-winners.html>.

<sup>182</sup> National Committee on Vital and Health Statistics Full Committee (Sep. 13, 2017), <https://www.ncvhs.hhs.gov/transcripts-minutes/transcript-of-the-september-13-2017-ncvhs-full-committee-meeting/>.

medicine practitioners, and urgent care facilities.).<sup>183</sup> A privacy advocacy group identified others not covered by HIPAA.

These include gyms, medical and fitness apps and devices not offered by covered entities, health websites not offered by covered entities, Internet search engines, life and casualty insurers, Medical Information Bureau, employers (but this one is complicated), worker's compensation insurers, banks, credit bureaus, credit card companies. many health researchers, National Institutes of Health, cosmetic medicine services, transit companies, hunting and fishing license agencies, occupational health clinics, fitness clubs, home testing laboratories, massage therapists, nutritional counselors, alternative medicine practitioners, disease advocacy groups, marketers of non-prescription health products and foods, and some urgent care facilities.<sup>184</sup>

Further, the current online advertising environment that developed over the last decade brings in another set of players who traffic in health information from time to time. This group includes advertisers, ad agencies, demand side platforms and supply side platforms, publishers, and others.<sup>185</sup> These companies and their functions are largely unknown to consumers.

In addition, another industry largely unknown to consumers includes data brokers and profilers that collect consumer information (including health information). The title of a May 2014 Federal Trade Commission report underscores the point about the invisibility of the data broker industry: *Data Brokers: A Call For Transparency and Accountability: A Report of the Federal Trade Commission*.<sup>186</sup>

How would a technology like blockchain apply to unregulated health information and offer protections to consumers? That is a difficult question to answer. Whether blockchain could actually and effectively provide consumers better control over third party uses of their personal information is an unaddressed matter. The commercial data environment involves considerable amounts of invisible or nonconsensual data collection and has little room for consumer involvement. Data collection can be the product of consumer activities and transactions; nontransparent tracking of consumer conduct on the Internet or otherwise; modeled data and data derived from algorithms; and in other ways. Many of the companies that collect this data would likely resist meaningful consumer involvement. Given the lack of regulatory authority over the industry, it is hard to envision a path that might lead to a privacy protecting blockchain application that would attract interest from the consumer data industry. The industry benefits

<sup>183</sup> NCVHS, Letter to the Secretary, Update to privacy laws and regulations required to accommodate NHIN data sharing practices (June 21, 2007), <https://www.ncvhs.hhs.gov/wp-content/uploads/2014/05/070621lt2.pdf>.

<sup>184</sup> World Privacy Forum, Patient's Guide to HIPAA, FAQ 9: Which Health Care Entities Must Comply With HIPAA?, <https://www.worldprivacyforum.org/2013/09/hipaaguide9-2/>.

<sup>185</sup> See, e.g., Ad Ops Insider, How RTB [Real Time Bidding] Ad Serving Works (2010), <http://www.adopsinsider.com/ad-serving/diagramming-the-ssp-dsp-and-rtb-redirect-path/>.

<sup>186</sup> <https://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf>. An earlier FTC report has more information on the data broker industry. See Protecting Consumer Privacy in an Era of Rapid Change: Recommendations for Businesses and Policymakers (2012), <https://www.ftc.gov/sites/default/files/documents/reports/federal-trade-commission-report-protecting-consumer-privacy-era-rapid-change-recommendations/120326privacyreport.pdf>.

today from the lack of consumer notice, involvement, and control. From the industry's perspective, blockchain's ability to give data subject some role in the use of their information might be characterized as a solution to something that the industry does not consider to be a problem.

The general problem of finding new technologies that offer the promise of privacy protection and that could be adopted in the unregulated world of consumer data is difficult. Technology may be as much of a threat to personal privacy as a protection.

## **B. Technologies can spark technical controversies**

Differential privacy offers a way to use data that gives a strong (but not absolute) guarantee that the presence or absence of an individual in a dataset will not significantly affect the final output of an algorithm analyzing that dataset. Here is a more detailed description.

Differential privacy is a rigorous mathematical definition of privacy. In the simplest setting, consider an algorithm that analyzes a dataset and computes statistics about it (such as the data's mean, variance, median, mode, etc.). Such an algorithm is said to be differentially private if by looking at the output, one cannot tell whether any individual's data was included in the original dataset or not. In other words, the guarantee of a differentially private algorithm is that its behavior hardly changes when a single individual joins or leaves the dataset -- anything the algorithm might output on a database containing some individual's information is almost as likely to have come from a database without that individual's information. Most notably, this guarantee holds for any individual and any dataset. Therefore, regardless of how eccentric any single individual's details are, and regardless of the details of anyone else in the database, the guarantee of differential privacy still holds. This gives a formal guarantee that individual-level information about participants in the database is not leaked.<sup>187</sup>

In some applications involving health data, differential privacy has utility. For example, one identified application is cohort identification, an activity that involves querying a patient database to identify potential recruits for a clinical trial.<sup>188</sup> There are more applications in health and other areas. Evaluating the technology or its possible health applications is not the point here.

The actual degree of privacy protection when using differential privacy depends on choices made when constructing the data set.<sup>189</sup> There are tradeoffs between privacy and accuracy. One

---

<sup>187</sup> Harvard University Privacy Tools Project, Differential Privacy, <https://privacytools.seas.harvard.edu/differential-privacy>.

<sup>188</sup> Fida K. Dankar, and Khaled El Emam, Practicing Differential Privacy in Health Care: A Review (2013), 5 Transactions on Data Privacy 35, 57 (2013), <http://www.tdp.cat/issues11/tdp.a129a13.pdf>.

<sup>189</sup> For a more rigorous discussion, see Thomas Steinke and Jonathan Ullman, Between Pure and Approximate Differential Privacy, 7 Journal of Privacy and Confidentiality 3 (2016), [https://privacytools.seas.harvard.edu/files/privacytools/files/between\\_pure\\_and\\_approximate\\_differential\\_privacy.pdf](https://privacytools.seas.harvard.edu/files/privacytools/files/between_pure_and_approximate_differential_privacy.pdf).

researcher describes how *differentially private* does not actually mean private and that it may be necessary to know more than the label.

Given recent publicity around differential privacy, we may soon see differential privacy incorporated as part of commercial data masking or data anonymization solutions. It is important to remember that "differentially private" doesn't always mean "actually private" and to make sure you understand what your data anonymization vendor is really offering.<sup>190</sup>

This begins to hint at the controversy. While differential privacy receives attention and increasing use, critics argue that “differential privacy will usually produce either very wrong research results or very useless privacy protections.”<sup>191</sup> These critics have their critics. One wrote of the just quoted article that it “has a long procession of factually inaccurate statements, reflecting what appears to be the authors' fundamental lack of familiarity with probability and statistics.”<sup>192</sup> These disagreements among experts can be impossible for policy makers to resolve and can make it difficult to make choices about the use of new technologies.

Outside of the HIPAA world, differential privacy has uses that may address some consumer privacy concerns, and disputes arise here as well. Apple uses differential privacy to allow it to mine user data while protecting user privacy. Researchers question the implementation of differential privacy that Apple chose to use and whether it protects privacy as well as it could.<sup>193</sup> Apple disputes the researcher’s conclusions.<sup>194</sup> Most Apple customers probably cannot evaluate the controversy and decide if Apple’s protections are adequate.

The challenge of evaluating technologies is not impossible in all cases. It may take time and, perhaps, consensus standards before we can agree on evaluating differential privacy applications for their actual degree of privacy protection. In the unregulated world of consumer data, it may be difficult to explain differential privacy to consumers no matter the context. It is also noteworthy that differential privacy at best provides a degree of privacy protections for limited applications. It does not address all aspects of privacy. One can easily foresee some companies promoting and misrepresenting their use of differential privacy or other technologies as protecting the interests of consumers. Consumers may have no way to evaluate these claims, which may be the point of the promotion.

From a policy perspective, a lesson is that it can be hard to evaluate technologies quickly or without expert assistance. Whether there would be any significant use of differential privacy in

---

<sup>190</sup> Paul Francis, What does it mean to be differentially private? (2017), <https://iapp.org/news/a/what-does-it-mean-to-be-differentially-private/>.

<sup>191</sup> Jane Bambauer, Krishnamurthy Muralidhar, and Rathindra Sarathy, Fool's Gold: an Illustrated Critique of Differential Privacy 16 Vanderbilt J. Ent. & Tech. 701(2014), <http://www.jetlaw.org/journal-archives/volume-16/volume-16-issue-4/fools-gold-an-illustrated-critique-of-differential-polic/>.

<sup>192</sup> Frank McSherry, Blog, Differential privacy for dummies (2017), <https://github.com/frankmcsherry/blog/blob/master/posts/2016-02-03.md>.

<sup>193</sup> Jun Tang, Aleksandra Korolova, Xueqiang Wang, Xiaofeng Wang, and Xiaolong Bai, Privacy Loss in Apple’s Implementation of Differential Privacy on MacOS 10.12 (2017), <https://arxiv.org/pdf/1709.02753.pdf>.

<sup>194</sup> See Andy Greenberg, How One of Apple's Key Privacy Safeguards Falls Short, Wired (Sep. 15, 2017), <https://www.wired.com/story/apple-differential-privacy-shortcomings/>.

the commercial arena remains to be seen, and it is equally unknown whether those applications would truly help data subjects. A mechanism that would require or encourage greater use in that arena of differential privacy or other new technology is not readily apparent in the current environment.

The same analysis may apply in roughly the same way to any other new technology that seeks to protect privacy. Encryption promotes privacy, but there are many controversies about the adequacy of encryption methods that mere mortals cannot evaluate. The next section offers an example.

### **C. Using technology to hide data linkage**

The value of encryption for protecting privacy and security is too well established to need explanation.<sup>195</sup> There is much to debate on the actual value of any given encryption method and on the implementation of the technology. As with the differential privacy, some of the more technical encryption issues come down to battles among experts.

In the unregulated world of consumer data, encryption in the form of hashing (a cryptographic function that masks the information being hashed) can be used to protect privacy and allow “de-identified” consumer profiling. A phone number properly hashed cannot be reconstructed from the output of the function. This can allow consumer data companies to claim that data is anonymous. In theory, a hashed record cannot be linked with other data.

Whether this use of encryption is meaningful can vary not so much on the strength of the hashing algorithm but on other factors. One analyst reports that different companies may use the same hashing function to anonymize a phone number or email address. The result is that two disparate records held by separate companies can be linked because each company’s hash produces an identical result. The effect is that “even though each of the tracking services involved might only know a part of someone’s profile information, companies can follow and interact with people at an individual level across services, platforms, and devices.”<sup>196</sup> Any claim that records are anonymized may be disingenuous at best.

A recent NCVHS letter to the Secretary acknowledged the risks of re-identification of data de-identified under the HIPAA Safe Harbor method.<sup>197</sup> The concern expressed in that letter was not precisely the same as discussed above, but the broad concerns are the same. Data supposedly de-identified can be subject to re-identification even when the methods used comply with official rules. In its recommendations, NCVHS pointed to non-technical means for bolstering de-identification.

**Recommendation 2: HHS should develop guidance to illustrate and reinforce how the range of mechanisms in the Privacy Rule, such as data sharing agreements,**

---

<sup>195</sup> Under the HIPAA security rule, encryption is an addressable technical safeguard. 45 C.F.R. § 164.312(a)(2)(iv).

<sup>196</sup> Wolfie Christl, *Corporate Surveillance in Everyday Life* (2017) (Cracked Labs), <http://crackedlabs.org/en/corporate-surveillance/#2/>

<sup>197</sup> NCVHS, *Recommendations on De-identification of Protected Health Information under HIPAA* (Feb. 23, 2017), <https://www.ncvhs.hhs.gov/recommendations-on-de-identification-of-protected-health-information-under-hipaa/>.

business associate agreements, consent and authorization practices, encryption, security, and breach detection, are used to bolster the management of de-identified data in the protection of privacy. Particular attention should be directed at the way in which business associate agreements should address obligations regarding de-identification and the management of de-identified datasets.<sup>198</sup>

There is no current process that requires unregulated users of health information to undertake any measures that would support effective use of de-identification methods. Further, for many commercial uses of health information, de-identification would undermine the value of the data to consumer data companies.

#### **D. Non-Technological Protections**

Technological methods for sharing data while protecting privacy have great promise. It is important; however, to remember that “old-fashioned” devices like data use agreements (DUAs), contracts, and the like can also support data sharing while keeping both those who disclose data and those who receive data accountable for their actions. Many examples of data use agreements exist, and the Centers for Medicare and Medicaid Services at HHS make widespread use of DUAs.<sup>199</sup> Under the Computer Matching and Privacy Protection Act of 1988 (an amendment to the Privacy Act of 1974), computer matching agreements regulate the sharing of personal information among federal agencies and between federal agencies and states.<sup>200</sup> The National Academy of Science report on innovations in federal statistics includes a discussion of administrative measures that protect data while protecting privacy.<sup>201</sup>

It is a given that there are many uncertainties and controversies about the scope of privacy as a public policy concern and a lack of clear consensus about the best ways to balance privacy against other interests. It is not necessary to resolve all of these conflicts at once. Processes that address issues in a formal way can be valuable. So DUAs usefully define rights and obligations in narrow areas. While technology holds much promise, it is not the only source of solutions.

In theory, non-technological protections might have application to unregulated health information. However, the motivation to use these measures may be absent with those seeking to make commercial use of health information.

---

<sup>198</sup> Id. at page 13.

<sup>199</sup> See, e.g., <https://www.cms.gov/Research-Statistics-Data-and-Systems/Files-for-Order/Data-Disclosures-Data-Agreements/States.html>, and <https://www.cms.gov/Research-Statistics-Data-and-Systems/Files-for-Order/Data-Disclosures-Data-Agreements/Enterprise-Privacy-Policy-Engine.html>.

<sup>200</sup> 5 U.S.C. § 552a(o).

<sup>201</sup> National Academy of Sciences, INNOVATIONS IN FEDERAL STATISTICS Combining Data Sources While Protecting Privacy 82-89 (2017), <https://www.nap.edu/catalog/24652/innovations-in-federal-statistics-combining-data-sources-while-protecting-privacy>.

## VI. Evolving Consumer Attitudes

Fairly assessing consumer attitudes on privacy and on health information privacy is hard. Consumers say one thing about privacy but their actions may not be consistent with their stated opinions. This is sometimes called the privacy paradox, and it is a common starting point in discussions about consumers and privacy. The privacy paradox is that consumers say that they are concerned about privacy, but their actions suggest that they do not actually care that much about privacy in the end. Discussions about the privacy paradox often focus more on online behavior, possibly because evidence and experiments are more easily collected in the online environment.

Experiments to measure consumer attitudes and actions do not report consistent results. Polls can show whatever result the pollster chooses to achieve. Industry points to lack of use of privacy tools by consumers, but the tools can be difficult to find, use, and maintain. The problem is not made any easier by a lack of agreement on just what privacy means.

Another common observation is that that consumers favor convenience over privacy. Consumer acceptance and use of social media is evidence that they happily share some personal information with others and do not act to stop monitoring of their activities. On the other hand, how consumers use social media suggests that many users actively control what they do and what they disclose even while they share other information.<sup>202</sup>

While not a direct measure of consumer concern about privacy, the use of ad blockers on the Internet continues to increase. A 2016 assessment found that almost 70 million American using ad blockers on desktops or laptops, and over 20 million using blockers on mobile devices.<sup>203</sup> Further increases appear likely.

Experiments measuring consumer behavior appear to obtain conflicting results. One experiment involving shopping showed that consumers tended to purchase from online retailers who better protected privacy, including paying a premium for privacy protections.<sup>204</sup> A different experiment involving MIT students and a Bitcoin wallet sought to measure whether the students wanted to disclose contact details of friends; whether they wanted to maximize the privacy of their transactions from the public, a commercial intermediary, or the government; and whether they would take additional actions to protect transaction privacy when using Bitcoin. The results

---

<sup>202</sup> It is sometimes said that teens do not care about privacy because of their intensive use of social media. A leading researcher on teens, Danah Boyd, Principal Researcher at Microsoft Research, writes about how teens understand and approach privacy. See, e.g., Danah Boyd & Alice Marwick, *Social Privacy in Networked Publics: Teens' Attitudes, Practices, and Strategies* (2011) ("There's a widespread myth that American teenagers don't care about privacy"), <http://www.danah.org/papers/2011/SocialPrivacyPLSC-Draft.pdf>.

<sup>203</sup> eMarketer, *US Ad Blocking to Jump by Double Digits This Year* (June 21, 2016), <https://www.emarketer.com/Article/US-Ad-Blocking-Jump-by-Double-Digits-This-Year/1014111>. See also Mark Scott, *Use of Ad-Blocking Software Rises by 30% Worldwide*, *New York Times*, Jan. 31, 2017, <https://www.nytimes.com/2017/01/31/technology/ad-blocking-internet.html>.

<sup>204</sup> Janice Y. Tsai, Serge Egelman, Lorrie Cranor, & Alessandro Acquisti, *The Effect of Online Privacy Information on Purchasing Behavior: An Experimental Study*, 22 *Information Systems Research* 254 (2011), <http://www.guanotronic.com/~serge/papers/isr10.pdf>.

found that the participants at multiple points in the process made choices inconsistent with their stated preferences.<sup>205</sup>

These are just two of many experiments. The second offers additional insights. The authors suggest that 1) small incentives (e.g., pizza for students) may explain why people who say they care about privacy relinquish private data easily; 2) small navigation costs have a tangible effect on how privacy-protective consumers' choices are often in contrast with stated preferences about privacy; 3) the introduction of irrelevant, but reassuring information about privacy protection makes consumers less likely to avoid surveillance, regardless of their stated preferences towards privacy.<sup>206</sup> Other research supports the last point. A poll conducted a few years ago found that consumers overvalue the presence of a website privacy policy and assume that websites with a “privacy policy” have strong, default rules to protect personal data.<sup>207</sup>

Another relevant aspect of conflicting consumer desires relates to the tradeoff between privacy and personalization. Polling suggests that consumer want both privacy and personalization.<sup>208</sup> Cass Sunstein, a well-known legal scholar, writes about the benefits and consequences, concluding that trusted choice architects can produce better balanced results.

In principle, personalized default rules could be designed for every individual in the relevant population. Collection of the information that would allow accurate personalization might be burdensome and expensive, and might also raise serious questions about privacy. But at least when choice architects can be trusted, personalized default rules offer most (not all) of the advantages of active choosing without the disadvantages.<sup>209</sup>

On the other hand, Shoshana Zuboff, Harvard University Berkman Center, views what she calls “surveillance capitalism” with more concern.

It is constituted by unexpected and often illegible mechanisms of extraction, commodification, and control that effectively exile persons from their own behavior while producing new markets of behavioral prediction and modification. Surveillance capitalism challenges democratic norms and departs in key ways from the centuries long evolution of market capitalism.<sup>210</sup>

---

<sup>205</sup> Susan Athey, Christian Catalini, & Catherine Tucker, *The Digital Privacy Paradox: Small Money, Small Costs*, Small Talk (2017) (MIT Sloan Research Paper No. 5196-17), [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2916489](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2916489).

<sup>206</sup> *Id.*

<sup>207</sup> Chris Jay Hoofnagle & Jennifer King, *Consumer Information Sharing: Where the Sun Still Don't Shine* (2008) (“Unfortunately, most consumers are under the misimpression that a company with a “privacy policy” is barred from selling data.”), [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=1137990](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1137990).

<sup>208</sup> See, e.g., Andrew Jones, *VentureBeat*, *Consumers want privacy ... yet demand personalization* (July 14, 2015), <https://venturebeat.com/2015/07/14/consumers-want-privacy-yet-demand-personalization/>.

<sup>209</sup> Cass R. Sunstein, *Impersonal Default Rules vs. Active Choices vs. Personalized Default Rules: A Triptych* (2013), [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2171343](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2171343).

<sup>210</sup> Shoshana Zuboff, *Big Other: Surveillance Capitalism and the Prospects of an Information Civilization*, 30 *Journal of Information Technology* 75 (2015), [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2594754](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2594754).

One obvious concern here is that better personalization may need *more* PII. At the bottom of the personalization slippery slope is a justification for the collection and use of every available scrap of data about every individual in order to provide more nuanced products, services, and advertising. The trail here also leads to concerns about discrimination, stereotyping, and personalized pricing. These debates, while important, are beyond the scope of this report.

Another general concern is that information collected by private sector data controllers then becomes available to governments, private litigants, and others for making decisions about the data subject far removed from the original purpose of collection. For example, it is possible that a consumer's transactions and activities could lead to the addition of the consumer's name to a No-Fly-List. Overall, the "proper" balance between personalization and privacy remains a very open subject.

Privacy management is another aspect of privacy that now receives some attention. From a consumer perspective, privacy management refers to the time and effort needed to read and understand privacy notices and to exercise choices about use and disclosure of PII by third parties when choices are available. For an average consumer, it may be somewhat challenging to read and understand the ever-changing privacy policies for a website like Facebook. However, doing the same for the dozens or hundreds or thousands of websites that collect PII about the consumer is impossible. Many companies that collect PII are invisible and unknown to consumers. Lorrie Cranor, a professor in the School of Computer Science and the Engineering and Public Policy Department at Carnegie Mellon University, calculated almost ten years that it would take an average consumer approximately 244 hours a year to read all relevant privacy policies.<sup>211</sup> If recalculated today, that number would be higher. The challenges of managing privacy may be a contributing factor to Joseph Turow's finding cited above that Americans are resigned to the lack of control over data and feel powerless to stop its exploitation.

In the end, it is unclear what type of privacy protections consumers want and what actions consumers will take to protect their privacy. There is evidence to be found on all sides.

Looking more narrowly at health privacy, results still show ambiguity. Are Americans concerned about health privacy? In a 2005 poll, 67% were somewhat or very concerned about the privacy of their personal medical records.<sup>212</sup> In a 2014 poll, 16 percent of respondents have privacy concerns regarding health records held by their health insurer; 14 percent have concerns about records held by their hospital; 11 percent with records held by their physician; and 10 percent with records held by their employer.<sup>213</sup> It is difficult to reconcile these two results, although the wording of the questions may account for the disparity.

While many showed relative indifference to privacy in the 2014 poll, in another poll consumers seemed to have a different view when it came to sharing records for research. When asked about

---

<sup>211</sup> Aleecia M. McDonald & Lorrie Faith Cranor, The Cost of Reading Privacy Policies, 4 I/S: A Journal of Law and Policy for the Information Society 543, 560 (2008-09), <http://lorrie.cranor.org/pubs/readingPolicyCost-authorDraft.pdf>.

<sup>212</sup> California Health Care Foundation, National Consumer Health Privacy Survey 2005, <http://www.chcf.org/publications/2005/11/national-consumer-health-privacy-survey-2005>.

<sup>213</sup> Truven Health Analytics, Data Privacy Health Poll (2014), [http://truvenhealth.com/Portals/0/NPR-Truven-Health-Poll/NPRPulseDataPrivacy\\_Nov2014.pdf](http://truvenhealth.com/Portals/0/NPR-Truven-Health-Poll/NPRPulseDataPrivacy_Nov2014.pdf).

willingness to share your health information with healthcare researchers anonymously, 53% said yes and 46 percent said no. The substantial minority that would not agree to *anonymous* data sharing seems hard to square with the general lack of concern about privacy of health records, although the two different questions asked about different aspects of health privacy.

Over the years, pollsters conducted many other polls about consumer views on health privacy. A 2009 Institute of Medicine report summarizes much of the polling.<sup>214</sup> One interesting point that appears to be consistent in polls is that while patients support health research, a majority prefers to be consulted before their information is available for research. This is true even if researchers receive no identifying information.<sup>215</sup> Thus, a headline that reports “Most Americans Would Share Health Data for Research” may be misleading because the poll only measured willingness to share anonymous records.<sup>216</sup> For health research that requires longitudinal linking of patient records over time and place, the availability of only anonymous records makes the research difficult or impossible.

There are other vaguer measures of popular interest in privacy. One measure comes from the media, where privacy stories of all types are commonplace. Privacy breaches and other more dramatic threats to privacy certainly attract media attention, although breaches are so ordinary that only large ones seem newsworthy. The 2017 Equifax data breach that affected more than 100 million individuals received long and detailed coverage in the press<sup>217</sup> as well as responses from legislators.

Another measure of popular interest is the passage of privacy legislation at the state level. In the two decades, a large number of privacy bills passed in the states, even if federal action was minimal. In particular, data breaches and the increased incidence of identity theft contributed to the passage of privacy legislation in most states.

On the other side, other privacy-affecting activities of companies appear to generate lesser degrees of attention, and consumer responses vary from occasional active reaction to more common indifference. The inconsistency of consumer response seems to be true for many routine consumer data activities, including unregulated health data. The lack of response may be due to ignorance.

In the end, it seems to be the case that anyone can find at least some support for any conclusion on consumer attitudes toward privacy.

#####

---

<sup>214</sup> Institute of Medicine, *Beyond the HIPAA Privacy Rule: Enhancing Privacy Improving Health Through Research* at chapter 2, (2009), <https://www.nap.edu/read/12458/chapter/1#iv>.

<sup>215</sup> *Id.* at 83.

<sup>216</sup> See National Public Radio, *Poll: Most Americans Would Share Health Data for Research* (Jan. 9, 2015), <http://www.npr.org/sections/health-shots/2015/01/09/375621393/poll-most-americans-would-share-health-data-for-research>.

<sup>217</sup> See, e.g., Brian Fung, *Equifax finally responds to swirling concerns over consumers’ legal rights*, *Washington Post*, Sep. 10, 2017, [https://www.washingtonpost.com/news/the-switch/wp/2017/09/08/what-to-know-before-you-check-equifaxs-data-breach-website/?utm\\_term=.678375a285db](https://www.washingtonpost.com/news/the-switch/wp/2017/09/08/what-to-know-before-you-check-equifaxs-data-breach-website/?utm_term=.678375a285db).