HHS Public Access

Author manuscript

Science. Author manuscript; available in PMC 2019 August 01.

Published in final edited form as:

Science. 2019 February 01; 363(6426): 448–450. doi:10.1126/science.aav5133.

Shadow Health Records Meet New Data Privacy Laws

W. Nicholson Price II[1,2,*], **Margot E. Kaminski**[3,4], **Timo Minssen**[2,5], and **Kayte Spector-Bagdady**[6,7]

[1]University of Michigan Law School, Ann Arbor, MI.

[2]Centre for Advanced Studies in Biomedical Innovation Law, Copenhagen, Denmark.

[3]University of Colorado Law School, Boulder, CO.

[4]Silicon Flatirons Center, University of Colorado, Boulder, CO.

[5]University of Copenhagen Faculty of Law, Copenhagen, DK.

[6]Department of Obstetrics and Gynecology, University of Michigan Medical School, Ann Arbor, MI.

[7]Center for Bioethics and Social Sciences in Medicine, University of Michigan Medical School, Ann Arbor, MI.

Abstract

Large sets of health data can enable innovation and quality measurement but can also create technical challenges and privacy risks. When entities such as health plans and health care providers handle personal health information, they are often subject to data privacy regulation. But amid a flood of new forms of health data, some third parties have figured out ways to avoid some data privacy laws, developing what we call “shadow health records”—collections of health data outside the health system that provide detailed pictures of individual health—that allow both innovative research and commercial targeting despite data privacy rules. Now that space for regulatory arbitrage is changing. The long arms of Europe’s new General Data Protection Regulation (GDPR) and California’s new Consumer Privacy Act (CCPA) will reach shadow health records in many companies. In this article, we lay out the contours of the GDPR’s and CCPA’s impact on shadow health records and health data more broadly, highlight critical remaining uncertainty, and call for increased clarity from lawmakers and industry on the use of such data for research.

BENEFITS, BARRIERS, WORKAROUNDS

With potential inputs such as genetic data, health history, lifestyle, phenotypes, and health outcomes, big health data has the capacity to translate information into knowledge more efficiently than ever. The current regulatory structure in the United States, however, was built neither to protect nor to enable such uses of big data to drive research and improve health. Although the United States does not have a general-purpose federal data privacy law,

*Correspondence to: wnp@umich.edu.

its most sweeping data privacy regulations—the Privacy and Security Rules of the Health Insurance Portability and Accountability Act (HIPAA)—target personal health information and aim to protect health data privacy (1). The Privacy Rule governs when protected health information (PHI) may be used or disclosed by health care providers, plans, or clearinghouses ("covered entities"). HIPAA allows use or disclosure purposes such as quality improvement but does not create a wholesale exception for research (2). Under regulations for federally funded research with human subjects (the "Common Rule"), researchers must generally get consent from subjects or obtain an Institutional Review Board waiver to use identifiable data. Thus, neither clinical nor research health information privacy structures were designed to accommodate rapid analysis of massive data (3).

At least partially to avoid these legal hurdles, some firms have developed workarounds to gather data about health. First, healthcare data are not the only health-related data that we generate. Fitness trackers, web searches, shopping histories, social media posts, and mobile personal health apps can reveal information about an individual's health (4). Consumer genetic services provide health information to the consumer and store health data (also relevant to blood relatives) for potential future research; a recently announced collaboration between 23andMe and GlaxoSmithKline (and a $300 million equity investment) demonstrates the scope of interest in using such data for research (5).

A second major workaround lets companies extract data from HIPAA or other health care data regimes (6). HIPAA is custodian-specific, regulating action of covered entities and business associates not the data themselves or whoever holds them (7). Covered entities can deidentify PHI, rendering it unprotected by HIPAA. They can then sell those data to brokers, who can link them with existing, non–HIPAA-protected data from outside health contexts, sometimes including identifiable information (6). Or patients can download their own medical records and share them. Once data are outside HIPAA's covered-entity protection, they can be used or reused by any actor who gets access (6).

Combining these sources can create shadow health records: less-regulated records about individuals with the same sort of information as standard health records—sometimes the same exact information—supplemented with data from other sources. Collections of shadow health records can be immense and key components of research. IQVIA, formerly IMS Health, one of the world's largest data brokers, has claimed to have "approximately 400 million comprehensive, longitudinal, anonymous patient records" and to know "85% of the world's prescriptions by sales revenue"; these data were assembled from more than 100,000 sources (6).

Such records can lead to the exact privacy risks that HIPAA and the Common Rule were designed to obviate but remain outside their ambit. Those regulations didn't foresee the masses of health-related data, and linked health inferences, that would be available in the future, so what was once a fair attempt at comprehensive health data privacy regulation ended up leaving unintentional space for arbitrage. These shadow health records thus represent an uneasy solution to a problem; they can be used for innovation or care improvement—or, more problematically, to target advertisements or identify high-cost patients to avoid. But benefits come at the cost of avoiding data privacy regimes.

NEW DATA PRIVACY LAWS

Two recent data privacy laws fundamentally change the legal landscape for shadow health records. Both the European Union (EU) and California recently put in place robust data privacy regimes that apply to personal information, including health data. These regimes govern large jurisdictions and are likely, in practice, to change compliance requirements for entities within and beyond those jurisdictions.

Before these two regimes, not all shadow health records went unregulated in the United States. They faced a patchwork of laws. Sectoral and state privacy laws might apply, such as the Genetic Information Nondiscrimination Act (GINA) and state laws targeted at consent, discrimination, and genetic testing generally. The Federal Trade Commission and state attorneys general enforce aspects of data privacy as part of consumer protection. But these new data privacy regimes in the EU and California add protections—and close loopholes.

The GDPR went into effect in May 2018 (8, 9). With exceptions, it applies to “personal data” that are “processed” by a wide range of public or private entities, including companies likely to hold or generate shadow health records (10). “Processing” includes nearly anything a company would do to data, and “personal data” includes health data. Physical or physiological data can fall under the GDPR even if they do not directly identify a person but rather make that person identifiable and even if they do not otherwise implicate health status (10).

The GDPR requires companies to obtain personal data legally (e.g., with consent); to collect and process only as much data as necessary; usually to notify individuals when their data have been received; and much more (10). Processing “data concerning health” is, by default, prohibited (8). Processors must either obtain explicit consent or fall under various GDPR exceptions, including for medical treatment, the “public interest in the area of public health,” and scientific research (8). But the scope of these exceptions remains uncertain, especially for research, in part because permitted conduct depends on EU or member-state law that has not yet been enacted and which may lead to divergent requirements. The GDPR also permits member states to enact or maintain more stringent laws around genetic data, biometric data, or “data concerning health” (8).

The GDPR potentially governs many companies outside the EU (10). It reaches data processing aimed at EU persons, such as by offering them a product or service; monitoring the behavior of EU persons within the EU, even if done from abroad; and other types of processing about EU persons, even if done abroad (8).

At the same time that the GDPR has expanded the extraterritorial reach of EU data privacy law, the mechanisms that U.S. companies use to export EU persons’ data are under judicial scrutiny. This makes for a risky legal environment. The GDPR’s actual reach is debatable, but U.S. companies with a transatlantic presence or that know they are monitoring EU persons will likely attempt to comply, given the risk of hefty fines (9).

For U.S. companies that don’t process EU persons’ data, a potentially more relevant new law is the CCPA, effective in 2020 (11). The CCPA applies only to personal information about

California residents. However, because of the size and economic influence of the state, many companies not based in California process information about California residents. As with the GDPR, “personal information” is defined very broadly (11, 12). Although the CCPA will not apply to PHI collected by entities already subject to HIPAA, it will apply to a wide variety of information in shadow health records (11).

The CCPA applies principally to larger businesses and data brokers, which excludes some smaller health data players. It creates notice and access requirements for businesses that collect, sell, or disclose information, and consumers may request that certain information be deleted and may opt out of the sale of their information (14). The CCPA lacks many GDPR protections, but may actually create additional or at least different requirements (12). Like the GDPR, the CCPA has exceptions for research; for instance, the right to deletion does not apply to consented “public or peer-reviewed scientific… research”—but the exception requires that the research be in “the public interest,” and the applicability of this exception to research by commercial entities remains unclear.

OPPORTUNITIES AND OBSTACLES

The value of GDPR and CCPA restrictions on data used to create shadow health records depends on context. Many have applauded the GDPR and the CCPA for prioritizing individual privacy and patient rights by establishing strict requirements and the need for organizational compliance (1). Those who regard shadow health records primarily as an impingement on privacy will likely welcome the GDPR and CCPA as closing—or at least shrinking—problematic loopholes.

The GDPR especially targets third-party information-processing companies that have gone unregulated in the United States by, for example, imposing notice requirements and other data privacy rights. By contrast, others might highlight the importance of shadow health records as a useful workaround for industry to promote health-care innovation and research in a kludgy system and would stress the negative impacts of heightened compliance thresholds on health system modernization and quality improvement.

On the positive side, these two laws could foster a new competitive environment in which health care providers and companies will compete on capability, procedures, and recruitment strategies to engage those whose data they seek. The GDPR and CCPA aim to let individuals control their data in a clear, transparent, and easy way. Studies show that people are often quite comfortable sharing their health information with corporate actors (and are sometimes even more willing to share than with health care providers) and that very few read detailed terms of use (13).

If these new data privacy laws enable competition in big-data research in a way that affirmatively protects individuals’ privacy and autonomy, that is progress. These laws might also level the playing field by making it harder for industry to skirt HIPAA restrictions that are imposed on academic medical research centers. On the negative side, additional hurdles —such as notifying individuals and gaining affirmative consent for sensitive-data processing —may exacerbate differences in innovative capacity between big players in the health and

life sciences (or big data competitors such as Google and Amazon) and smaller firms that lack resources to ensure compliance.

Although these regimes are territorial, their impacts are likely to be broader. The laws of a large territory can force companies toward higher standards of compliance elsewhere. It can be less expensive for a company to adjust its behavior broadly than to create multiple standards for multiple jurisdictions or risk substantial liability; multiple global companies have already updated their privacy policies worldwide to comply with GDPR requirements. Nevertheless, if the GDPR and CCPA impose too many barriers or too much uncertainty, data brokers might shift away from health data or refocus their efforts on less-regulated jurisdictions.

Additional guidance—particularly, clarification regarding research exemptions for work done in the “public interest”—will be critical under both the GDPR and CCPA. This clarity is necessary to avoid chilling the potential innovative impact of the health data industry while ensuring that enough protections exist so that the individuals that make up big health data are knowing and willing participants. For example, the GDPR refers to exceptions for “scientific research,” the “public interest,” and “public health” without clearly defining these overlapping terms or addressing dual-use endeavors. Although GDPR guidance suggests that “scientific research” should be defined broadly and include both technological development and privately funded research (Recital 159), it elsewhere suggests that public health and public interest exceptions “should not result in personal data being processed for other purposes by third parties…” (Recital 54). Even where the GDPR permits research exceptions, it requires “appropriate safeguards” to protect individual privacy rights—without clarifying what those safeguards must be [for example, in Articles 89(1) and 9 and Recitals 52 and 54].

The European Data Protection Board should issue updated guidelines on health data to clarify the scope of these exceptions and required safeguards. In the absence of formal EU-wide guidelines, individual member states could enact clarifying laws that keep public health research in mind or, less formally and with less effort, direct their data protection authorities to issue guidelines or informal guidance. The United Kingdom’s data protection authority, the Information Commissioner’s Office, has been effective at issuing both more formal guidelines and informal online FAQs [for example, (14)].

The CCPA provides a more detailed definition of “research” and more detailed safeguard-like requirements than those of the GDPR, but there is room for clarification [§ 1798.140(s)]. The CCPA limits its research exception to deidentified information not used for a commercial purpose and used “solely for research purposes that are compatible with the context in which the personal information was collected.” The law defines “commercial purposes” broadly [§ 1798.140(f)]; guidance from the state’s attorney general on how much this covers commercially funded research would be helpful. Also helpful would be guidance on what exactly constitutes a research purpose “compatible with the context” in which information was gathered.

At least in the context of the GDPR, industry could play an important role in developing best practices. By coming together to suggest transparent codes of conduct for collecting, storing, processing, and sharing data, the health data sector could influence how EU regulators understand the balance between individual rights and the public interest in this space (Article 40). Two early examples of industry attempting this include the Privacy Best Practices for Consumer Genetic Testing Services (15) and the CARIN Code of Conduct (16). Especially as requirements and research exceptions remain unclear, such codes of conduct, if transparent and comprehensive, could help develop a workable balance between privacy and big-data innovation concerns to guide regulators and legislators working to clarify the contours of the GDPR and CCPA.

Last, if the current U.S. Congress turns to federal data privacy law, as some have predicted, it would do well to explicitly consider the value and challenge of big data research on health in particular, avoiding the arbitrage opportunities of HIPAA and clarifying the murky research exceptions of the GDPR and CCPA. Ultimately, both transparency and accountability are necessary to increase much-needed public trust in data science.

REFERENCES AND NOTES

1. Gostin LO, Halabi SF, Wilson K, JAMA 320, 2334 (2018).
2. Price WN, II, Minn. Law Rev. 102, 101 (2018).
3. Spector-Bagdady K, Shuman AG, Otolaryngol. Head Neck Surg. 158, 405 (2018). [PubMed: 29494320]
4. Tanner A, Our Bodies, Our Data: How Companies Make Billions Selling our Medical Records (Beacon Press, 2017).
5. Molteni M, WIRED 3 8 (2018); www.wired.com/story/23andme-glaxosmithklinepharma-deal.
6. Terry NP, Yale J Health Policy Law Ethics 17, 143 (2017).
7. Cohen IG, Mello MM, JAMA 320, 231 (2018). [PubMed: 29800120]
8. Regulation (EU) 2016/679 Of the European Parliament and of the Council of 27 April 2016 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).
9. Marelli L, Testa G, Science 360, 496 (2018). [PubMed: 29724945]
10. Hoofnagle CJ et al., SSRN Scholarly Paper ID 3254511, Rochester, NY, 2018; https://papers.ssrn.com/abstract_id=3254511.
11. California Consumer Privacy Act of 2018.
12. de la Torre L, GDPR matchup: The California Consumer Privacy Act 2018; https://iapp.org/news/a/gdpr-matchup-california-consumer-privacy-act.
13. Ostherr K et al., Big Data Soc. 4, 1 (2017). 10.1177/2053951717704673
14. Information Commisioner's Office, General Data Protection Regulation (GDPR) FAQs for small health sector bodies (2018); https://perma.cc/2TGT-RYUQ
15. https://perma.cc/7M4P-X964
16. https://perma.cc/3BH2-YRDB