DAY ONE PROJECT

Building Trust in the Health Data Ecosystem

Jennifer C. Goldsack

July 2020

Summary

Pending bipartisan "Cures 2.0" legislation is intended to safely and efficiently modernize healthcare delivery in the wake of the novel coronavirus (COVID-19) pandemic. Such modernization is contingent on access to high-quality data to power innovation and guided decision making. Yet over 80% of Americans feel that the potential risks of companies collecting their data outweigh the benefits.[1] To ensure the success of Cures 2.0, provisions must be added that bolster public trust in how health data are used.

Addressing the largely unregulated activities of data brokers—businesses that collect, sell, and/or license brokered personal information—offers a budget-neutral solution to the public's crisis of faith in privacy. Building a well-governed health-data ecosystem that the public can trust is essential to improving healthcare in the United States.

Challenge and Opportunity

Despite dedicating massive amounts of spending to treating chronic disease,[2] chronic-disease burdens in the United States are significantly worse than in other OECD countries.[3] One reason for this poor return on investment is the weakness of the U.S. health-data ecosystem. This weakness manifests in multiple ways:

- At a health clinic, it is challenging for medical providers to integrate data from patients' everyday lives with clinical data in patient medical records to optimize care.
- Millions of digital health records exist for catastrophic diagnoses like cancer, each providing vital information about what treatments are most likely to save lives. But because these data are trapped in silos instead of being at the fingertips of every treating physician, physicians must rely only on their best judgement to determine what treatments will work best for any given patient.
- Devastating diseases like Alzheimer's have no treatments and no cure. Successful development of therapies requires earlier diagnoses powered by better information, as well as better integration of data generated during routine clinical care with data from clinical trials.
- Much health or health-relevant data is inaccessible to researchers and others who could put it to use to address health crises like the COVID-19 pandemic. From return-to-work initiatives to targeted prevention efforts, the absence of a trusted data infrastructure renders it nearly impossible to use data-driven approaches in health crisis response.

[1] Auxier, B.; et al. (2019). Americans and Privacy: Concerned, Confused and Feeling Lack of Control Over Their Personal Information. Pew Research Center. Available at https://www.pewresearch.org/internet/2019/11/15/americans-and-privacy-concerned-confused-and-feeling-lack-of-control-over-their-personal-information/.

[2] Anderson, G.; Horvath, J. (2004). The Growing Burden of Chronic Disease in America. *Public Health Reports*, 119: 263–270.

[3] Tikkanen, R.; Abrams, M.K. (2020). U.S. Health Care from a Global Perspective, 2019: Higher Spending, Worse Outcomes? The Commonwealth Fund. Available at https://www.commonwealthfund.org/publications/issue-briefs/2020/jan/us-health-care-global-perspective-2019.

Improving the health of all Americans in the 21st century requires a health-data ecosystem that allows authorized users to unlock data-driven insights to improve public health, while also protecting individuals from data-based discrimination.

A variety of programs and initiatives that leverage high-quality data for public health benefits are already underway. These include the $1.5 billion "All of Us" Research Program authorized by the 21st Century Cures Act, continued development of the Food and Drug Administration (FDA)'s Sentinel system, and a flurry of more recent attempts to integrate data from diverse sources as part of COVID-19 response strategies.

Cures 2.0 must build on and extend such efforts. Each of the legislation's six key focus areas require broad access to diverse, representative health data from a variety of sources to achieve the legislation's overall goal of safely and efficiently modernizing delivery of treatments to patients.[4]

Productively leveraging digital health data will not be possible if trust in how data are handled is so low that individuals and institutions refuse to share data. Incorporating pragmatic, achievable provisions designed to bolster public trust in the national health-data ecosystem is essential to the success of Cures 2.0.

Plan of Action

As U.S. Reps. DeGette (D-CO) and Upton (R-MI) continue to develop the bipartisan "Cures 2.0," baseline protections against digital health data discrimination must be included. Specifically, Cures 2.0 should include provisions that:

(1) Prohibit data brokers from selling data that report inferred health status and inferred health risk to third parties, for those parties to use to make adverse or discriminatory decisions against individuals and populations.

(2) Require the Federal Trade Commission (FTC) to maintain a list of data brokers who aggregate and sell information on American citizens and residents, including information about any data breaches by these companies.

(3) Require data brokers to provide opportunities for consumers to modify incorrect information and/or opt out of data collection.

(4) Levy a tax on net income from data brokers' sale of aggregated data, earmarking tax revenue to support FTC enforcement of—and other activities related to—the above provisions.

[4] The focus areas are (i) pandemic health and preparedness; (ii) caregiver integration; (iii) patient engagement in healthcare decision making; (iv) diversity in clinical trials; (v) Food and Drug Administration (FDA) modernization, focusing on digital health technologies and the use of real-world evidence; and (vi) Centers for Medicare and Medicaid Services (CMS) modernization, focusing on coverage and reimbursement.

A key objective of these provisions is to prevent the sale of health information inferred from consolidated data collected by data brokers to inform adverse or discriminatory decisions. This in turn will protect individuals and populations by default from harms based on assumptions about their current health and future health risk.

Absent such provisions, it is possible that a data broker could infer (for instance) an individual's mental-health status from aggregated social media, GPS, spending, and internet-search data. The broker could then sell or license that information to institutions such as insurers, employers, lenders, and schools—institutions that could use the information to engage in harmful adverse and discriminatory practices (e.g., higher insurance rates or application denial).[5]

The provisions stated above will not provide comprehensive privacy and data protection to individuals, nor are they intended to. Rather, these provisions would provide immediate and baseline protections for individuals by addressing the vulnerabilities posed by largely unregulated data brokers.

These budget-neutral provisions are intended to build public trust in how individuals' data may be used relative to their health. Adopting the provisions would represent an important step to improving healthcare in the United States through the development of a useful, high-quality, and well-governed health-data ecosystem.

[5] In the context of this memo, "adverse and discriminatory practices" generally refers to practices that deny access or add barriers to access (for example, increased costs and/or documentation requirements) to services and opportunities based on inferred health status and risk.

Frequently Asked Questions

Why are existing health privacy laws such as HIPAA inadequate?

HIPAA has a very limited scope: it applies only to data created or held by healthcare providers, health insurers and plans, and a tricky category of "data clearinghouses". HIPAA was never intended to be—and is not—a comprehensive health-privacy law. In addition, the limited protections it offers have not kept pace with the rapidly evolving and expanding scope of digital health data.

Why target data brokers?

Data brokers are businesses that collect, sell, and/or license the personal information of consumers—consumers that the brokers do not have a direct relationship with—to third parties. Data brokers have been involved in many high-profile data scandals that have recently undermined public trust in privacy and data protections, such as the Cambridge Analytica scandal.[6]

Yet with the exception of limited state regulations in Vermont and California,[7] data brokers remain largely unregulated in the United States. There are no controls around how data brokers derive, infer, and predict health status for individuals. The risk of harm from inaccurate health assumptions add to the harms of discrimination based on health status.

Even when individuals technically consent to their data being shared with data brokers (for example, by agreeing to the contract for a store loyalty card or by accepting the terms of service of an app), those individuals are almost certainly not in control of who gets to see the data aggregated about them by the broker, what inferences are made, or who this information is sold to, or for what use. With 93% of adults saying that it is important to be in control of who can get information about them, limiting the capabilities of data brokers to infer and sell health information is foundational to regaining public trust in the digital health-data ecosystem. [8]

Why should these provisions be incorporated into Cures 2.0?

Cures 2.0 sets ambitious goals of safely and efficiently modernizing delivery of care to patients. These goals cannot be achieved without public trust in the health data ecosystem.

[6] The Cambridge Analytica Story, Explained. WIRED. Available at: https://www.wired.com/amp-stories/cambridge-analytica-explainer/.

[7] Vermont General Assembly. (2018). H.764 (Act 171): An act relating to data brokers and consumer protection. Available at https://legislature.vermont.gov/bill/status/2018/H.764; California State Legislature. (2019). Assembly Bill (AB) 1202: Privacy: Data brokers. Available at https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201920200AB1202.

[8] Madden, M.; Rainie, L. (2015). Americans' Attitudes About Privacy, Security and Surveillance. Pew Research Center. Available at https://www.pewresearch.org/internet/2015/05/20/americans-attitudes-about-privacy-security-and-surveillance/.

The limited provisions proposed for Cures 2.0 in this memo are intended to provide immediate support for the broader success of Cures 2.0 by reinforcing trust in the health-data ecosystem, while acknowledging that more comprehensive legislation on health-data rights is still necessary.

These provisions are intended to be foundational and complementary to other pending pieces of legislation, including approaches to health-data privacy during the COVID-19 pandemic from both sides of the aisle,[9,] bipartisan legislation focused on data collected from health tracking devices and apps as well as DNA testing kits,[10] and even calls to define privacy as a human right.[11]

What are examples of similar policies and initiatives that have been successful?

The Genetic Information Nondiscrimination Act (GINA), passed in 2008, protects Americans against discrimination based on their genetic information when it comes to health insurance and employment.[12] This law has advanced personalized medicine and improved health for individuals without fear of discrimination. Similar, but more broadly applied protections for individuals' digital specimens would advance health through data reuse while reducing the risk of harm from data misuse.

In 2018, Vermont enacted the first data-broker privacy law in the United States, requiring that data brokers disclose to individuals which data are being collected and to permit individuals to opt out of collection.[3] In 2019, California passed AB 1202, requiring data brokers to register with and provide certain information to the state attorney general.[13] A bill requiring data brokers to register with the Secretary of State is also pending in Illinois.[14] These state laws are proving popular. Passing similar legislation at the federal level would not only extend data-privacy benefits to all American citizens and residents but would also avoid a patchwork of state laws that could make compliance more onerous for businesses and data access more difficult for researchers and other authorized users.

[9] U.S. Congress. (2020). COVID-19 Consumer Data Protection Act of 2020. Available at https://www.congress.gov/bill/116th-congress/senate-bill/3663/text?r=83&s=3; U.S. Congress. (2020). Public Health Emergency Privacy Act. Available at https://delbene.house.gov/uploadedfiles/public_health_emergency_privacy_act_-_as_introduced.pdf.

[10] U.S. Congress. (2020). Protecting Personal Health Data Act. Available at https://src.bna.com/I7O.

[11] Cappello, L. (2019). Surveillance is a fact of life, so make privacy a human right
Surveillance is a fact of life, so make privacy a human right. *The Economist*. Available at https://www.economist.com/open-future/2019/12/13/surveillance-is-a-fact-of-life-so-make-privacy-a-human-right.

[12] Equal Opportunity Employment Commission. (2016). Genetic Information Nondiscrimination Act. 29 CFR Part 1635.

[13] California State Legislature. (2019). Assembly Bill (AB) 1202: Privacy: Data brokers.

[14] Illinois General Assembly. (2019). House Bill (HB) 2871: Data Broker Registration Act. Available at http://www.ilga.gov/legislation/fulltext.asp?DocName=&SessionId=108&GA=101&DocTypeId=HB&DocNum=2871&GAID=15&LegID=119175&SpecSess=&Session=.

About the Author[15]

Jennifer C. Goldsack co-founded and serves as the Executive Director of the Digital Medicine Society (DiMe), a 501(c)(3) non-profit organization dedicated to advancing digital medicine to optimize human health. Jen's research focuses on applied approaches to the safe, effective, and equitable use of digital technologies to improve health, healthcare, and health research. She is a member of the Roundtable on Genomics and Precision Health at the National Academies of Science, Engineering and Medicine. Jen holds a master's degree in chemistry from the University of Oxford, England, a masters in the history and sociology of medicine from the University of Pennsylvania, and an MBA from the George Washington University.

About the Day One Project

The Day One Project is dedicated to democratizing the policymaking process by working with new and expert voices across the science and technology community, helping to develop actionable policies that can improve the lives of all Americans, and readying them for Day One of a future presidential term. For more about the Day One Project, visit dayoneproject.org.

[15] Jennifer gratefully acknowledges the support of the Day One Project accelerator and the input and constructive comments from the following experts during the development of this memo:

- Deven McGraw, Chief Regulatory Officer at Citizen and former Deputy Director for Health Information Privacy at the Office for Civil Rights of the U.S. Department of Health and Human Services.
- Dena Mendelsohn, Director of Health Policy and Data Governance at Elektra Labs and former Senior Policy Counsel for Consumer Reports.
- Mona Siddiqui, Senior Vice President for Strategy and Quality, Enterprise Clinical Management at Humana and former HHS Office of the CTO Chief Data Officer.