


consumers about health information privacy or security.

Imperfect, HIPAA has had its critics.<sup>4</sup> Implementing it has been expensive for the health care industry. It did not allow wronged patients to bring civil lawsuits and did not completely preempt state health privacy laws. Exceptions allow third parties to access health information for legal proceedings, public health activities, and biomedical research about which the public remains poorly informed. Covered entities may err on the side of overcompliance, so that medical workers excessively fear making punishable errors, while patients are inconvenienced and needlessly denied access to

 An audio interview with Dr. Allen is available at NEJM.org

information. And of course, HIPAA does not safeguard physical or decisional privacy related to health care or make care more respectful to and affordable by all.

Yet HIPAA regulations critically aim to balance privacy protection with promotion of information access and technologies to improve health care quality and efficiency. Frankly, regulatory momentum, along with popular culture, has been pulling toward

greater data sharing and less privacy. Yet privacy is a kind of power; without it, health care consumers are at the mercy of those who would control, exploit, and manipulate our data. Big business and algorithms have greatly diminished our ability to exercise meaningful control over our data privacy.

The Covid-19 pandemic has revealed the extent to which our technology infrastructure allows employers and public health officials, for better or worse, to track, trace, and monitor people's symptoms, illnesses, and contacts. HIPAA regulations may be an institutional headache, but medical identity theft, ransomware attacks, data breaches, weak encryption, de-anonymization risks, wearable devices generating sensitive data, big data analytics, and discrimination are bigger headaches. Strong, well-informed regulations, with periodic revisions, can continue making a positive difference.

Privacy lawyers' assessments of HIPAA's impact skew positive — a perspective not universally shared by a health care industry saddled with the compliance burden. On HIPAA's 10th birthday, attorney Daniel Solove noted that

HIPAA had not bankrupted health care, shut down research, and paralyzed industry, as critics had feared. Instead, it “paved the way to real benefits for consumers through greater access to quality care.”<sup>5</sup> At 25, HIPAA is further along in paving the same important road.

Disclosure forms provided by the author are available at NEJM.org.

From the University of Pennsylvania School of Law, Philadelphia.

This article was published on June 5, 2021, at NEJM.org.

1. Goldstein MM, Pewen WF. The HIPAA Omnibus Rule: implications for public health policy and practice. *Public Health Rep* 2013; 128:554-8.
2. HHS proposes modifications to the HIPAA Privacy Rule to empower patients, improve coordinated care, and reduce regulatory burdens. Department of Health and Human Services, December 10, 2020 (<https://www.hhs.gov/about/news/2020/12/10/hhs-proposes-modifications-hipaa-privacy-rule-empower-patients-improve-coordinated-care-reduce-regulatory-burdens.html>).
3. Yaraghi N, Gopal RD. The role of HIPAA Omnibus Rules in reducing the frequency of medical data breaches: insights from an empirical study. *Milbank Q* 2018;96:144-66.
4. Hall MA. The HIPAA headache. *Hastings Cent Rep* 2008;38:7-8.
5. Solove DJ. HIPAA turns 10. *J AHIMA* 2013;84:22-8.

DOI: 10.1056/NEJMp2100900

Copyright © 2021 Massachusetts Medical Society.

## HIPAA and the Leak of “Deidentified” EHR Data

Kenneth D. Mandl, M.D., M.P.H., and Eric D. Perakslis, Ph.D.

The permissible sharing of patient data among health care organizations and their business associates for treatment, payment, and operations purposes has led to a torrent of electronic health record (EHR) data flowing out of health care provider silos. The Health Insurance Portability and Accountability Act (HIPAA) also permits business associates to deidentify data on behalf of a

health care provider, insurance plan, or clearinghouse (so-called covered entities); once data are deidentified, the business associate may use them freely, unless it is contractually prohibited from doing so. Organizations that don't qualify as business associates under HIPAA may also gain access to and use deidentified data sets. Such policies have enabled the rise of a multibillion-dollar industry

comprising dozens of health-data aggregation companies and hundreds more companies producing tools and technologies that aggregate, link, and monetize EHR data.

This phenomenon has been amplified by the explosion of data production since the Health Information Technology for Economic and Clinical Health (HITECH) Act of 2009 began promoting widespread adoption of EHRs to en-

able accurate and complete patient records, real-time learning, better-coordinated care, accelerated biomedical discovery, and exchange of digital data directly with patients. It is ironic that although patients (and their physicians) still have difficulty obtaining complete medical record information in a timely fashion, the HIPAA Privacy Rule permits massive troves of patients' digital health data to traverse the medical-industrial complex unmonitored and unregulated.

Privacy is essential for reducing the potential for abuse of power, supporting self-determination and individual preferences, and allowing people to preserve their reputations and avoid stigma. Although the HIPAA Privacy Rule governs uses of identifiable data, it doesn't apply to data that are considered deidentified, either as determined by experts or under the “safe harbor method,” which requires removal of 18 specific identifiers (such as name, address, and date of birth).

On the one hand, assembling vast data sets can help support a public good — the development of a learning health system, in which data that are routinely collected during care delivery continuously drive ever-more-intelligent treatment decisions. But markets for secondary use of patient data don't always serve the interests of patients or the public. For example, a data-aggregation company might target physicians and patients for pharmaceutical detailing, which could drive up drug prices and result in overprescribing.

Moreover, although the deidentification process is often treated as infallible, it is not. Nor is a particular method required for monitoring the success of deidentification efforts. Even after many deidentification-related pro-

cesses, individual patients can potentially be reidentified on the basis of only a handful of attributes.<sup>1</sup> Deidentification technologies relying on encryption could be vulnerable to future advances in computing. In the absence of contractual controls governing data produced by a covered entity and shared with a business associate, if something goes wrong, only patients are harmed; the United States doesn't have a comprehensive data-privacy law, and none of the various privacy-related laws or regulations protects patients from the potentially harmful use of deidentified data. There is no duty to report instances in which data have been reidentified or linked to external data sources, such as financial records, and patients have little or no opportunity for redress in cases of reidentification.<sup>2</sup>

Research involving deidentified data is generally conducted without institutional review board oversight. Furthermore, the existence of a free market for deidentified data can negatively affect data quality, especially since an analyst or researcher cannot go back to the original data source to remove duplicate data, properly validate data, or correct errors. These shortcomings can lead to the production of error-prone and nonreproducible research.<sup>3</sup>

As medical practice becomes more digitally driven, health care providers will increasingly face incentives to obtain access to multi-institutional data, since very few organizations have large enough patient populations to gather sufficient data to support even basic functions, such as diagnosis and decision support. When organization leaders see that they can achieve short-term gains by commercializing data, they often enter into data-sharing agree-

ments with third parties, with no benefits of data-driven knowledge accruing to the organization. Some health care organizations enter data-sharing agreements to acquire access to multi-institutional data mediated by commercial third parties. These aggregated data sets may become more expensive as health care delivery organizations increasingly depend on them to practice in a digital medicine ecosystem.<sup>4</sup> If an institution's goal is to make money, selling patient data might offer short-term advantages. If its goals are to foster the development of a learning health system, high-quality research, and entry into durable and transparent compacts with patients, other strategies are more promising.

Broadly speaking, two approaches could help address the torrential leak of deidentified health record data. One approach would be to **establish best practices for data protection among data providers**. The other would be to **strengthen legal and regulatory protections for patients**.

Health care institutions could better protect their own interests and patients' privacy by treating deidentified data more similarly to protected health information. First, institutions should inform patients — using consent documents and privacy notices — that their data may be used to support a learning health system and, when appropriate, may be shared with commercial parties. Visual and interactive media may be useful adjuncts to written consent documents.

Second, when data must be shared with external parties, **proper contractual controls should be implemented** to ensure that the data never pass beyond the users specified in the arrangement, that they cannot be linked with other

data sets without the permission of the original provider, and that reidentification is prohibited without the permission of the provider. This approach would also enable institutions to be fully transparent in disclosures to patients about how they are using and sharing their data.

Third, a promising method for permitting data use involves “behind the glass” access for outside parties so that data don’t leave the institution. A health care delivery organization working with a third-party service provider, researcher, or other collaborator can establish an enclave with a shared analytic workspace. When multiple institutions need to share data, federated systems can bring analytic tools to the data — which remain at the originator sites. When necessary, data can be combined on a project-by-project basis, under protective contracts or data-use agreements.

To improve legal and regulatory oversight, other states and the federal government could join California in making reidentification of deidentified health data illegal. Such policies could have an important effect on the market — but they might not prevent a malicious actor from exposing a patient’s medical information.

It’s also worth closely examin-

ing the advantages and drawbacks of the European Union’s General Data Protection Regulations (GDPR) “right to erasure” policy. This rule ensures that a data subject can choose to have their information erased from a data set without undue delay when the data are being used for purposes other than the original one, when the subject withdraws consent or objects to any use of the data, or when the data are being used unlawfully.<sup>5</sup> This rule ensures that data subjects can choose to have their information erased from a data set without undue delay when the data are being used for purposes other than the original one, when a subject withdraws consent or objects to use of the data, or when the data are being used unlawfully. If patients don’t opt out of data sharing at the beginning of a project, however, their records cannot be subsequently located in a truly deidentified data set. In crafting laws or regulations inspired by GDPR provisions, U.S. policymakers could preserve the ability to use data in positive ways. In a learning health system, for example, an opt-out model for deidentified data sets might bias the data sets and prevent accurate analysis.

HIPAA and its privacy rule

were crafted in the pre-EHR era. Health systems, legislators, and regulators now have an opportunity to protect health record data to a greater degree than the law mandates, while actively promoting and supporting the beneficial uses<sup>5</sup> of large data sets for improving health and optimizing health care delivery.

Disclosure forms provided by the authors are available at NEJM.org.

From the Computational Health Informatics Program, Boston Children’s Hospital, and the Department of Biomedical Informatics, Harvard Medical School — both in Boston (K.D.M.); and the Duke Clinical Research Institute, Duke University Medical Center, Durham, NC (E.D.P.).

This article was published on June 5, 2021, at NEJM.org.

1. Rocher L, Hendrickx JM, de Montjoye Y-A. Estimating the success of re-identifications in incomplete datasets using generative models. *Nat Commun* 2019;10:3069.
2. McGraw D, Mandl KD. Privacy protections to encourage use of health-relevant digital data in a learning health system. *NPJ Digit Med* 2021;4:2.
3. Mehra MR, Ruschitzka F, Patel AN. Retraction-hydroxychloroquine or chloroquine with or without a macrolide for treatment of COVID-19: a multinational registry analysis. *Lancet* 2020;395:1820.
4. Mandl KD, Bourgeois FT. The evolution of patient diagnosis: from art to digital data-driven science. *JAMA* 2017;318:1859-60.
5. Bovenberg J, Peloquin D, Bierer B, Barnes M, Knoppers BM. How to fix the GDPR’s frustration of global biomedical research. *Science* 2020;370:40-2.

DOI: 10.1056/NEJMp2102616

Copyright © 2021 Massachusetts Medical Society.

## What Did I Sign Up For?

Jaclyn Heilman, M.D.

To say that the Covid-19 pandemic has changed all our lives dramatically is an understatement. The population of North Philadelphia is no exception. In a city where gun violence was rampant before Covid, bloodshed has only increased with the upheaval of the structure of every-

day life. Over the past year, there have been more than 2200 shooting victims and 500 homicides in Philadelphia, reflecting increases of 54% and 40%, respectively, over 2019.<sup>1,2</sup> In November 2020, while completing a trauma rotation, I witnessed the aftermath of such violence. One boy, along with

many others from that time, will always haunt me.

When he arrived in the trauma bay, he was alive. He had been shot in the chest. I immediately started working to gain access to his femoral vein as another doctor monitored his airway. His legs were kicking as he screamed,