PROTECTING HEALTH DATA PRIVACY *and* IMPROVING PATIENT CARE

A Report of the Aspen Health Strategy Group

HEALTH STRATEGY GROUP
aspen institute

Foreword by Kathleen Sebelius and William Frist

Edited by Alan R. Weil, Alexandra J. Reichert, and Karyn Feiden

PROTECTING HEALTH DATA PRIVACY *and* IMPROVING PATIENT CARE

A Report of the Aspen Health Strategy Group

HEALTH STRATEGY GROUP
aspen institute

Foreword by Kathleen Sebelius and William Frist

Edited by Alan R. Weil, Alexandra J. Reichert, and Karyn Feiden

support for this report provided by:

Robert Wood Johnson Foundation

Google Health

Laurie M Tisch ILLUMINATION FUND

The mission of the **Aspen Health Strategy Group (AHSG),** part of the Health, Medicine & Society Program at the Aspen Institute, is to promote improvements in policy and practice by providing leadership on complex health issues. AHSG brings together senior leaders representing a mix of influential sectors, including health, business, philanthropy, and technology, to tackle a single health issue annually through year-long, in-depth study. Co-chairs are Kathleen Sebelius, 21st U.S. Secretary of Health and Human Services and former Governor of the State of Kansas, and William Frist, former U.S. Senator from Tennessee and former Senate Majority Leader.

The topic of AHSG's seventh annual report is protecting health data privacy and improving patient care. This compilation opens with a consensus report based on the group's in-depth learning process, followed by a set of background papers. Taken together, these papers explore the positive transformative power of health data, but also the many privacy challenges and concerns raised by "big data" collection, use, and analytics.

The Aspen Institute
2300 N Street, NW
Suite 700
Washington, DC 20037

Published in the United States of America in 2023 by The Aspen Institute

Printed in the United States of America
Publication Number: 23/002

HEALTH
STRATEGY GROUP
aspen institute

CO-CHAIRS

Kathleen Sebelius, 21st U.S. Secretary of Health and Human Services (2009-2014); former Governor, State of Kansas (2003-2009)

William Frist, former U.S. Senator (TN) (1994-2006); former U.S. Senate Majority Leader (2003-2007)

MEMBERS PARTICIPATING IN THIS REPORT

Richard Baron, President and CEO, American Board of Internal Medicine

Richard Besser, President and CEO, Robert Wood Johnson Foundation

Raphael W. Bostic, President and CEO, Federal Reserve Bank of Atlanta

Gail K. Boudreaux, President and CEO, Elevance Health

Dena Bravata, Healthcare Entrepreneur

Rosalind Brewer, CEO, Walgreens Boots Alliance

Toby Cosgrove, Executive Advisor and former CEO and President, Cleveland Clinic

Deborah DiSanzo, President, Best Buy Health

Victor Dzau, President, National Academy of Medicine

David Feinberg, Chairman, Oracle Health

Harvey Fineberg, President, Gordon and Betty Moore Foundation

Helene Gayle, President, Spelman College

Ai-Jen Poo, President, National Domestic Workers Alliance; Executive Director, Caring Across Generations

David J. Skorton, President and CEO, Association of American Medical Colleges

Antonia Villarruel, Dean, University of Pennsylvania School of Nursing

EX-OFFICIO MEMBERS

Alex M. Azar II, 24th U.S. Secretary of Health and Human Services (2018-2021)

Sylvia Mathews Burwell, President, American University, 22nd U.S. Secretary of Health and Human Services (2014-2017)

Michael O. Leavitt, Founder, Leavitt Partners; 20th U.S. Secretary of Health and Human Services (2005-2009)

Tommy G. Thompson, 19th U.S. Secretary of Health and Human Services (2001-2005); AHSG Co-chair (2015 – 2021)

Donna Shalala, 18th U.S. Secretary of Health and Human Services (1993-2001)

Louis W. Sullivan, Chair, Sullivan Alliance to Transform America's Health Professions, Association of Academic Health Centers; 17th U.S. Secretary of Health and Human Services (1989- 1993)

David Mathews, 11th U.S. Secretary of Health, Education and Welfare (1975-1977)

STAFF

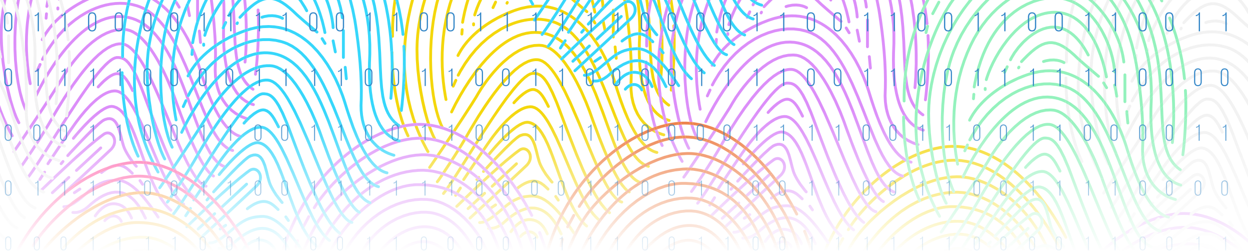
Alan R. Weil, Editor-in-Chief, *Health Affairs*

Alexandra Reichert, Ph.D. Student, Vanderbilt University

Ruth J. Katz, Vice President; Executive Director, Health, Medicine & Society Program; Director, Aspen Ideas Health, Aspen Institute

Katya Wanzer, Manager of Operations, Health, Medicine & Society Program, Aspen Institute

Raven Tucker, Project Coordinator, Health, Medicine & Society Program, Aspen Institute



December 2022

It is my great privilege to introduce the seventh annual report of the Aspen Health Strategy Group (AHSG).

Since its launch in 2015, AHSG has taken on some of America's most complex health challenges. By bringing together a diverse set of leaders—representing health systems, the private sector, professional associations, philanthropies, and universities—and diving deep into a single topic every year, AHSG helps to bring bold ideas forward. Past reports have explored end-of-life care, the opioid epidemic, chronic disease, antimicrobial resistance, maternal mortality, and the health harms of incarceration.

The focus of this report is equally pressing: health data privacy. The era of "big data" is upon us, fueled by advances in data collection, data mining, analytics, and computing power. Tremendous opportunities to personalize medicine, inform medical decisions, speed drug development, and much more reside in the available data. Yet, access to this vast body of information also poses challenges to the privacy rights of individuals and the optimal functioning of health systems.

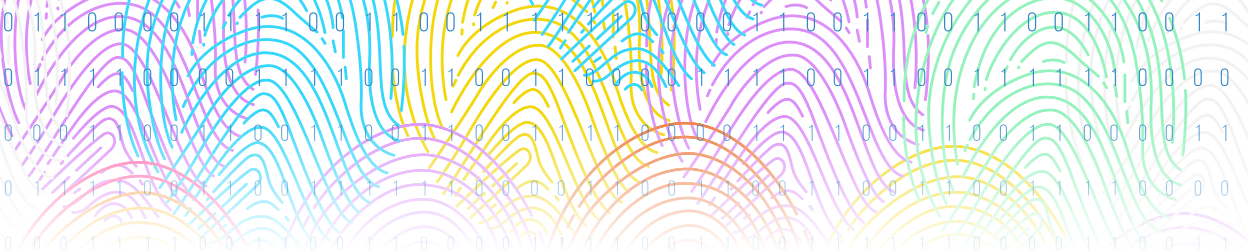
The Aspen Health Strategy Group, housed within the Aspen Institute's Health, Medicine & Society Program, is well positioned to consider how data can be used to their greatest advantage while also ensuring that appropriate privacy safeguards are in place. The personal and professional networks of a diverse membership give AHSG unique reach, helping to draw the attention of policymakers and other influencers to the group's recommendations and inspiring them to take action.

Kathleen Sebelius and William Frist, both long-time partners to the Aspen Institute, serve as AHSG cochairs. Kathleen Sebelius, a former U.S. Secretary of Health and Human Services and former Governor of the State of Kansas, has helped to lead AHSG since its inception. Bill Frist, former U.S. Senator from Tennessee and former Senate Majority Leader, became co-chair in 2020. I am grateful for the gift of their time and their contributions to the dialogue, leadership, and action that drive the Aspen Institute's mission to create a free, just, and equitable society.

My thanks, as well, to all members of the Aspen Health Strategy Group, and to you, our readers, whose interest and support gives our work meaning.

Daniel R. Porterfield

Dan Porterfield
President and CEO
Aspen Institute



Contents

Foreword

Kathleen Sebelius
AHSG Co-Chair

William Frist
AHSG Co-Chair

After two long years in which we could only meet virtually, members of the Aspen Health Strategy Group (AHSG) convened in Aspen, Colorado in June 2022. It was a tremendous pleasure to be with our colleagues again, enjoying the vigorous conversations that are only possible when all of us are in the same room. *Protecting Health Data Privacy and Improving Patient Care*, the seventh annual report of the Aspen Health Strategy Group, emerged from that gathering.

We came together well aware that data are transforming the health care landscape. Businesses, government, health systems, clinicians, and patients are all tapping into an extraordinary wealth of knowledge that offers so much hope for preventing, detecting, and treating disease. At the same time, vast data storehouses are introducing numerous legal, medical, and ethical issues into the equation. Patient autonomy remains a paramount obligation of health systems and providers but in the age of digital health, there is little consensus on how best to protect individual preferences. Numerous questions remain unanswered about what constitutes appropriate consent, how data can be legitimately commercialized, and how best to foster innovation and the robust infrastructure that can support data standardization and interoperable systems. Many of the laws and institutional policies that could guide decision-making are dated, inadequate, or non-existent.

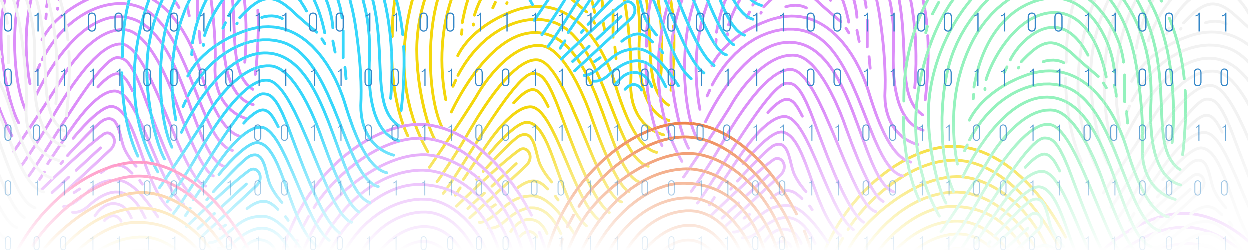
Our exploration of data privacy was informed by subject matter experts who developed the four background papers included in this report and then joined our convening to provide further insights. Two other presentations also enriched our conversation. Mollyann Brodie, who heads the Public Opinion and Survey Research program at the Kaiser Family Foundation, discussed public views of health care data and privacy while Helen Nissenbaum, professor of information science at Cornell Tech, focused on the ethical and political implications of digital technologies as data privacy policies are considered. Alan Weil, editor-in-chief of *Health Affairs*, ably moderated three days of deep discussions, as he has done at all of our sessions. He led the effort in synthesizing key themes from the discussion and capturing the group's five big ideas for overhauling health data privacy rules.

It has become standard practice at our convenings to hear from individuals who have been directly affected by the issues we are exploring. This year, we were joined by Jeri Lacks Whye and David Lacks Jr., the grandchildren of Henrietta Lacks, and by Rebecca Skloot, author of *The Immortal Life of Henrietta Lacks;* together, they put a human face on the importance of using personal medical information appropriately. The HeLa line of cancer cells, taken from Henrietta Lacks without her knowledge, has been used in research for decades and are the foundation for extraordinary advances in medicine. Yet they never should have been harvested without her consent.

The family of Henrietta Lacks and author, Rebecca Skloot

Through its leadership on complex health issues, AHSG remains committed to promoting improvements in policy and practice. Our work would not be possible without the generosity of our funders. The Robert Wood Johnson Foundation and the Laurie M. Tisch Illumination Fund have been steadfast supporters from the start and we are deeply grateful to them, and to Google, which provided funding for the first time this year. Importantly, we note that the framework and language of this report reflect the perspectives of the authors, but not necessarily the views of these funders.

On behalf of the Aspen Health Strategy Group, our thanks to everyone who made the 2022 program possible. We believe our work makes a genuine difference in shaping the public and private sector response to profoundly challenging health issues and recognize all of the logistical, conceptual, and scholarly efforts that must be invested to get it right.



ASPEN HEALTH STRATEGY GROUP REPORT

Five Big Ideas on Protecting Health Data Privacy and Improving Patient Care

Part 1

"We must redesign our approach to health data privacy to honor the ethical value of privacy and to earn public support for using health data to positive ends."

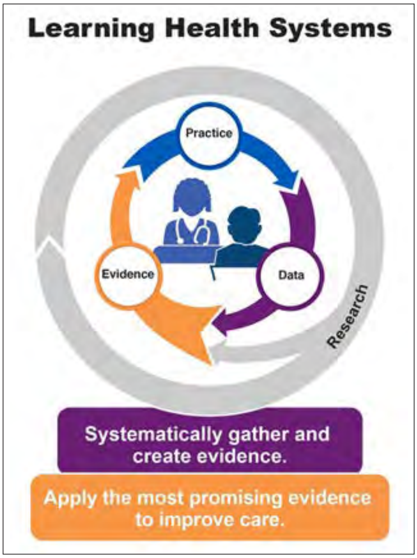
– THE ASPEN HEALTH STRATEGY GROUP

Five Big Ideas on Protecting Health Data Privacy and Improving Patient Care

We acknowledge and thank Jeri Lacks Whye and David Lacks Jr., grandchildren of Henrietta Lacks, for their participation in our meeting and for sharing their family's story.

Introduction

Dramatic growth in the collection and uses of health data raises two major questions: Are these data being put to good use? And, is enough being done to protect patient privacy? Some uses of health data are unambiguously good, such as when patients voluntarily consent to sharing data as they participate in clinical trials that will improve understanding of diseases and how to treat them. Other uses are more questionable, such as when a smartphone application gathers information about a person's health habits and sells it to a commercial enterprise. Some data collection and use fully respects patient privacy, as occurs when a person makes an informed judgment about sharing data, based on a clear explanation of what is being collected and how it will be used. In other instances, data collection and use may not respect patient privacy—for example, when consent is obtained simply by asking a person to check a box that allows for broad data collection, without understanding what will be collected and how it will be used, and with little opportunity to consider the implications of the choice.


Learning Health Systems
Practice
Evidence
Data
Research
Systematically gather and create evidence.
Apply the most promising evidence to improve care.

The National Academy of Medicine defines a "learning health care system" as one "that is designed to generate and apply the best evidence for the collaborative health care choices of each patient and provider; to drive the process of discovery as a natural outgrowth of patient care; and to ensure innovation, quality, safety, and value in health care" (Institute of Medicine 2007). Patients cared for in such a system realize profound benefits as their care reflects the accumulated experience and wisdom of all who have preceded them.

Data collection, use, and analysis are the foundations of a learning health care system. Yet, if patients and the public at large do not trust the entities that perform these functions, they will withhold their data. In addition to the ethical imperative to honor people's desire for privacy, the practical stakes for respecting preferences and enforcing legal protections related to health data privacy are very high.

When people discuss privacy, they often refer to balancing the individual right to privacy against other social or individual goals, such as achieving medical advances or enabling consumer convenience. While the desire for balance is a useful metaphor, it also has important limitations. Balance presumes one fulcrum point, but in a large, heterogeneous country with divergent values, different people place that point at different locations. How do we establish broadly applicable public policies that reflect a multiplicity of values? To the extent that we rely on individual choices to honor those values, can we feel confident that people have the information they need and a realistic opportunity to choose whether to share their data?

The Aspen Health Strategy Group (AHSG) selected health data privacy as its topic for discussion in 2022, its seventh year. This group of leaders within and outside health care spent three days considering the topic, with the assistance of subject matter experts. In addition to participation by the four authors whose papers are summarized below, the group benefited from presentations by Mollyann Brodie of the Kaiser Family

Foundation; Helen Nissenbaum of Cornell University; Jeri Lacks Whye and David Lacks Jr., grandchildren of Henrietta Lacks; and Rebecca Skloot, author of *The Immortal Life of Henrietta Lacks*—and by discussions with all of them. The group emerged with five big ideas to modernize the country's health data privacy rules.

AHSG's goal is to promote improvements in health policy and practice by providing leadership, ideas, and direction on important and complex health issues. Co-chaired by Kathleen Sebelius, former Governor of Kansas and former U.S. Secretary of Health and Human Services, and William Frist, a physician and former U.S. Senate Majority Leader, the group comprises senior leaders across sectors that include health, business, philanthropy, and technology. More information about the Aspen Health Strategy Group can be found on the Aspen Institute website (http:// www.aspeninstitute.org/aspen-health-strategy-group). This report captures the conversations of the group, but no specific section or statement in the report should be considered to represent the opinion of any individual member.

Background

Four experts prepared papers to support the AHSG in its discussions. The papers are published in full as part of this volume; brief summaries appear below. Publication of the papers and inclusion of the summaries in this report do not imply agreement by the AHSG members with their conclusions or recommendations.

In "Health Information Privacy in the Digital Health Age," Deven McGraw depicts the health data privacy landscape. Noting the lack of a single definition of health data, McGraw writes, "The relevance of a piece of data to describing a person's health depends more on how the data are used rather than the characteristics of the data themselves."

The Health Insurance Portability and Accountability Act of 1976 (HIPAA), McGraw notes, "is a sectoral law; it covers only certain types of entities and generally does not extend to organizations and businesses outside of the traditional health care ecosystem." Since HIPAA's coverage is defined by who is doing the collecting, rather than what is being collected and for what purpose, a large amount of health-relevant data is outside HIPAA protection. Examples include data collected by "social media platforms, health and wellness apps, smartphones, life insurers, retailers, credit card companies, and internet search engines." The Federal Trade Commission Act (FTCA) governs privacy in many contexts, including most businesses, but it relies upon each company's own commitments to privacy and does not create substantive standards for privacy protection.

Even as there are gaps in health data privacy provisions, McGraw explains that the U.S. health sector is "plagued by a lack of other data to inform optimal clinical care." Public policy focuses on privacy, consent, and deidentification, not the beneficial applications of health data that could improve patient outcomes and population health. A more effective strategy, McGraw argues, "incorporates protections and stimulates, encourages, or even demands responsible use as a mechanism to propel digital medicine initiatives in the United States."

While several states have enacted new data protections, and Congress is considering action as well, McGraw notes that these new provisions, which are not specific to health data, place too much reliance upon notice and consent, shifting the burden of protecting privacy to the individual. Pending legislative action, McGraw argues for voluntary leadership on privacy protections by the health sector itself.

Anita Allen notes the long history of respect for health data privacy and defends its ongoing importance in "Health Data Privacy in the Balance: Evolving Values and Priorities." Referencing legal cases from the 19th century to the present, Allen finds that health privacy protections have served three policy goals: "enforcing customary morality, constraining public policy in the interest of individual rights, and governing a complex health infrastructure."

Until the early 2000s, social norms favored privacy. Allen states, "Dignity, self-determination, and well-being require opportunities for privacy and private choices, making privacy an aspect of the common good. Because it facilitates wellness and independent thought and action, privacy is vital for democratic life."

The pendulum has now swung toward data sharing as the path to better health. Today, "Under a new narrative, privacy interests are lightly addressed through data security; practices of informed consent, notice, and opt-out rights; and trust and transparency measures." Allen notes, "Much is at stake in the shift from health data privacy to health data disclosure as a dominant norm and preference."

The preferred metaphor for privacy and data sharing is "balance," but Allen notes the limitations of this term, as "one person's balance is another person's skew." Citing

Obasagie and Darnovsky, Allen states, “The traditional emphasis in academic and policy discussions of privacy, informed consent, and choice grounded in individual freedom can neglect the ways in which health disparities and structural injustices contribute to poor health, constrain choice, and reduce opportunity.” Allen closes by reminding us, “The challenge ahead is to figure out how to give both health innovation and privacy their due.”

In “Deidentification to Enhance Health Data Sharing,” Bradley Malin notes, “Clinical care generates a large amount of data related to an individual’s health that can serve as the basis for biomedical research” and “[d]eidentification is a mechanism that was developed to make it easier to share health data.”

Deidentification is a process that transforms personal data into a format that neither directly identifies nor includes information that can be used to identify an individual. Once data are deidentified, they are no longer covered by HIPAA. HIPAA provides two mechanisms for deidentification: safe harbor, in which certain identifying information is removed from the data, and expert determination, which allows for someone with “appropriate knowledge” of statistical methods in deidentification to apply generally accepted principles to determine that the risk of reidentification is very small. Malin also describes the HIPAA provisions for creating a limited data set, which is not fully deidentified, but is stripped of patient identifiers while retaining information that can be particularly useful for population health analysis and public health surveillance.

Malin notes a number of benefits associated with data deidentification. For example, given the difficulties associated with obtaining patient consent, datasets compiled from people who have given consent to use their data are not representative of the social and demographic characteristics of the population as a whole. This form of bias

can be reduced or eliminated in deidentified data, where consent is not needed. Deidentified data can also readily be combined with other types of data to create a more complete picture of the population's health.

Malin also describes deficiencies associated with the deidentification approach. In particular, improved data analytics present a growing risk of reidentification. In addition, since deidentified data do not fall under HIPAA, unauthorized releases of data need not be reported and there are no consequences when they occur. Indeed, deidentified data can be bought and sold without patients having any knowledge about it, which raises ethical concerns regarding data privacy and individual autonomy.

"New investments and agreement on a robust data infrastructure that supports data standardization and interoperability are required if we are to achieve desired changes in health system design and practice," writes Kenneth Mandl in "The Value and Uses of Health Data in the Clinical Ecosystem." Mandl highlights the benefits of a "learning health system," which continuously uses patient data to inform care for each future patient.

Such a system depends upon the efficient exchange of information across providers and sites of care. Mandl states, "There are emerging data exchange regimes, new technologies producing interoperable systems, novel governance models for intelligent data use across sites of care, and emerging business models for data aggregation. Each of these advances has implications for patient autonomy, privacy, and protection from harm."

There are many ways to facilitate data exchange. Mandl describes options that include patients controlling and sharing their own data, commercial aggregation of large datasets, research networks, and data sharing consortia. Central to the functioning of any of these mechanisms is interoperability—the ability to move data easily and simply across systems. As part of the 21st Century Cures Act, rules regarding interoperability will go into effect at the end of 2022. As Mandl points out, "New organizations are emerging to take advantage of regulated interoperability."

Mandl closes by asserting, "If the goal is a learning health system with standardized data sets yielding improved care and insights into disease causes and treatments, it is incumbent upon us to monitor the ecosystem, continually refining regulatory and legal frameworks and behavioral expectations for health systems, third-party apps, and companies aggregating and commercializing data."

Framing the Issue

Five themes emerged in the group's discussion that helped guide the development of this year's big ideas.*

- **The health data ecosystem is vast and growing rapidly**

  We are in the midst of an explosion in the quantity and uses of health data that is invisible to most people and likely to accelerate into the future. This growth is occurring along multiple dimensions: how data are collected, who is collecting data, what data are collected, and how data are used.

  When people referred to health data at the time HIPAA was enacted, they primarily meant paper charts that included handwritten notes made by clinicians during an office visit or hospital stay, along with physical copies of laboratory and imaging results.

  Within the health care sector, the nearly universal adoption of electronic health records (EHRs) has increased both the availability of data and its volume. EHRs prompt clinicians to complete predefined forms so that far more data are collected at each patient encounter.

* Unless noted otherwise, the data in this report come from presentations to the group or the background papers prepared by subject-matter experts and published in conjunction with this report.

High-fidelity images and laboratory results are stored electronically. Genomic sequencing can now be done at low cost, generating an entirely new type of health data that is stored in the patient's EHR. Millions of blood and tissue samples reside in databanks at health centers around the world.

With growing attention to the social determinants of health, medical records sometimes include information related to a patient's social needs, such as language spoken, housing stability, food security, and the caretaking capabilities of other residents in the patient's home.

Citing earlier work with Kenneth Mandl, McGraw introduces the concept of health-relevant data. Such data begin with the traditional health data collected and used by the health care system but also include information produced by consumer wearables, consumer-facing apps, and internet searches, as well as what a person buys at the grocery store, demographic data related to where they live, and more. Using this broader conception, health-relevant data are now collected continuously by a tremendous variety of companies and organizations. Insurance companies and administrative personnel in multiple clinical and nonclinical settings collect health-relevant data, as do stores and online retailers, and such collection is part of the eligibility determination processes of myriad federal, state, and local social programs. These data are gathered both actively, with the data subject knowingly providing them to a third party, and passively, through interactions with technology that the person views as entirely separate from their health.

The explosion of data collection, analytics, and uses has placed health data in the hands of numerous commercial enterprises, ranging from small startups to large technology companies, many of which are not regulated under HIPAA. The business models for these enterprises are highly varied as are their methods of data collection and use. The ability to monetize health data is growing rapidly, a trend that is likely to continue.

Not long ago, the primary uses of a patient's medical record were review by a clinician prior to a visit and recording of notes and care plans for later reference. Knowledge of the effects of treatments was acquired largely through clinical trials.

The combination of the proliferation of electronic records, a broader understanding of the factors affecting people's health, and massive increases in data storage and computational power has yielded entirely new uses for health data. Just a few examples are algorithms designed to improve care plans, predictive models to suggest candidates for care management or early intervention, and risk assessments used to adjust payment levels.

All signs suggest these trends will continue and expand, largely outside the view of individual patients.

- **Current rules fail to protect health data privacy adequately**

HIPAA and the FTCA provide critical protections for health data, and they continue to serve essential purposes. However, the dramatic expansion of the type, quantity, and uses of health data described above leaves significant gaps in the data privacy regime. A few limitations stand out.

HIPAA focuses exclusively on data held by covered entities, which are health system actors and specific entities that interact with them. HIPAA has specific rules to protect the privacy and security of patient data and carries significant penalties for unauthorized use or disclosures of that data. Yet, except when they obtain the data through a formal business associate agreement, many organizations and institutions that possess and use health data operate outside of HIPAA's purview. This includes large technology firms such as Apple, Alphabet (owner of Google), Meta (owner of Facebook), and many companies running consumer-facing applications (apps). HIPAA also focuses on how data are stored, shared, and used, with few provisions related to how, by whom, or when data are collected.

The FTCA is directed at a particular sector—commercial enterprises—and while it requires organizations to comply with their own policies, it does not create any substantive standards. Thus, many organizations are excluded from the FTCA's provisions, and even those that are included have wide latitude in their actions.

There is growing understanding that obtaining consent as the basis for permitting reuse of data is an inadequate process for protecting privacy. Consent forms are difficult to understand and are often presented to patients at a time or place where it is impossible to consider their implications. People agree to data privacy policies without reading them. The proliferation of health data, health data collectors, and health data uses has undermined much of the protection that consent rules provided in the past.

Allowing unlimited use of data that have been deidentified, as HIPAA does, fails to meet today's reality of data mining, data matching, and computational power. The growing number of entities holding data, combined with the complete deregulation of data that has been deidentified using standards developed in a different era, means the opportunities for reidentification have grown and will continue to do so.

While a handful of states have adopted new data privacy laws, this does not fully address current challenges. Health data travel across state lines, limiting the effectiveness and viability of a state-by-state approach. These new laws are not specific to health, so they fail to reflect the unique benefits of sharing information in the health context. Since they are quite new, it is not yet possible to determine the overall effects of these laws. And, of course, only a few states have taken action, which leaves the vast majority of Americans operating outside their provisions.

- **Safeguarding privacy is essential to realizing the benefits of health data use**

There are tremendous health benefits to maximizing the positive uses of health data. A learning health care system requires incorporating the lessons of today's patients into the care that tomorrow's patients will receive. The potential benefits of robust use of health data are particularly great for people with less common conditions, where data from millions of patients may be needed to find the small number who have the condition.

Similarly, the potential for supplementing clinical trial data with real-world evidence to identify effects of drugs, devices, and procedures on the population as a whole, and on specific subpopulations, can only be realized through large, connected health datasets.

All this requires people to be willing to share their health data. But weak or inadequate privacy protections, or the perception of weakness, undermines this willingness. Public concerns about data collection and use are substantial and widespread. People generally have little understanding of the current health data privacy regime; they express low trust in the existing system while supporting more government regulation.

Much can be learned from a small number of high-profile data breaches or uses of data that were inconsistent with people's general expectations. Each of these has yielded significant backlash, which is part of what is driving the push in Congress and among states to update data privacy laws in general—not simply those related to health. To the public at large, uses of data that fall outside of norms and expectations are perceived as a violation of the social contract, regardless of whether people clicked "accept" on a company's privacy policy.

Actual protection of health data privacy and the perception that the health data environment honors individuals' sense of fairness are both necessary to create an environment where health data can be collected and used for the benefit of individual patients and the public at large.

- **Health data use creates benefits and harms that are inequitably distributed**

  While health data have many beneficial uses, the costs of health data misuse do not fall equally. Health data have been used to harm specific populations, most notably racial and ethnic minorities, women, and the LGBTQ+ community.

Much of the public has heard of the U.S. government's intentional withholding of beneficial syphilis treatment for Black men in Tuskegee, Alabama, in the name of research. While the story of Henrietta Lacks, whose tissue was taken without her family's knowledge or consent and used for experimentation around the world, is familiar to many, we learned from her family about subsequent invasions of their privacy and exclusion from important decisions about the use of her tissue decades later. These are just two examples that represent dozens more with shared elements: the medical establishment experimenting on the bodies of Black Americans without consent, or with consent obtained through deception.

The right to health privacy, which many take for granted, has long been denied to poor people and people of color. Medical information must be shared to apply for a host of public benefits that are unrelated to health care. Before the nondiscrimination provisions of the Affordable Care Act were implemented, medical information was routinely used to deny people health insurance coverage. Current concerns that medical and nonmedical data will be used in legal actions related to newly imposed abortion restrictions are only the most recent example of health data being used to oppress women and those who treat them.

The benefits of medical treatments that derive from analyzing health data are also inequitably distributed. Most emerging technologies, whether drugs, devices, or new diagnostic tools, are expensive and may be financially out of reach for people without good health insurance or the ability to pay on their own. Public insurance programs, particularly Medicaid, have at times been slow to cover emerging treatments that are costly.

Inequities are multiplied when they affect both data collection and data use. For example, there is well-documented overrepresentation of people of European descent in genomic databases, while people from the rest of the world are underrepresented. Similarly, people of color are underrepresented in clinical trials. These biases can be traced in part to a history of exclusion and mistrust, as well as to current data collection practices that result in lower participation levels by historically excluded groups. Inequities reduce the value and precision of advances in diagnosis and treatment for underrepresented groups. For example, real-world evidence is increasingly used to refine our understanding of a drug's efficacy, but if that evidence does not draw from a fully representative population, our understanding of efficacy for different subgroups will be limited.

- **Rules regarding health data privacy should consider the type of data and how and by whom the data will be used**

  People demonstrate every day that they are willing to trade off a certain amount of privacy to achieve other goals, such as convenience. Even those most protective of their own privacy may be willing to share their health data with a trusted clinician in order to obtain an accurate diagnosis and effective treatment. To the individual, the right to privacy is rarely viewed as absolute.

Medical Record
Your Personal
Name
Date of birth
Gender
Nationality
Unit No.
Street No.
Street
Town/City/Suburb
State
Post Code
Country
DEPLOYMENT
Yes
No
1 Communicate purpose, vision and values to employees
2 Communicate purpose, vision and values to customers, supplier/partners and other external parties

While myriad types of data are health-relevant, a person's views of the privacy protections needed are not necessarily identical across all such data. For example, a person likely feels more strongly about the need for privacy regarding their medical records than they do about a record of the food they buy at the grocery store. They may be more worried about internet search records that tie to their sexual behavior than they are to a search for influenza. Even as we acknowledge the growing collection and uses of health-relevant data, our laws need to differentiate between core health data and broader health-relevant data.

Other key factors in how the public views data sharing reflect knowledge of who holds the data and the intended uses. Sharing data with a trusted organization in order to improve an individual's own medical care, or that of others, is viewed very differently than sharing data with a distant corporation that will sell it to others to enable more targeted marketing of products.

In a large and heterogeneous country with a rapidly changing health data ecosystem, drawing clear boundaries around what all people will and will not accept is impossible, yet the inability to draw uniform boundaries does not mean none should exist. A robust data privacy ecosystem will provide clarity as to who holds data and how they will be used, and it will enable people to elect whether to share that data based on their understanding of these factors.

Five Big Ideas to Protect Health Data Privacy and Improve Patient Care

Changes in the health data privacy rules of the United States are needed if we are to respect privacy and achieve the potential of using those data appropriately. The Aspen Health Strategy Group offers five big ideas to do so.

1. Congress should update federal health data privacy laws

While HIPAA and the FTCA serve important purposes, Congress should update federal health data privacy laws to reflect current uses and practices. The updated health data privacy regime should:

- attach to the data being protected and apply regardless of what entity or enterprise holds the data;
- eliminate some of the distinctions that currently exist in HIPAA and the FTCA;
- include provisions that reflect current data analytic capacity when it comes to deidentifying (and the potential for reidentifying) data; and
- establish ongoing regulation of deidentified data, including prohibition of reidentification.

2. Health data privacy laws should reflect social norms

Health data privacy policy and practice should be based on principles that reflect a combination of social and individual values. Heterogeneity in the importance that people place on privacy does not imply the absence of social norms that should be reflected in law.

Federal law should:

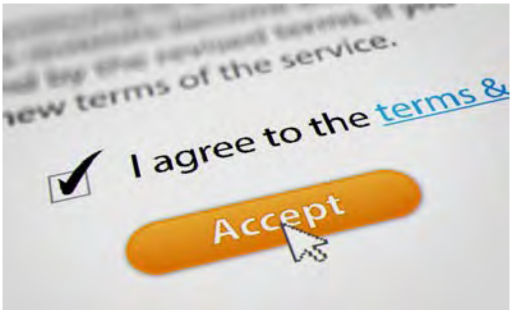
- Prohibit certain data collection practices and data uses that fall outside the reasonable expectations that a typical patient or consumer would have regarding what and when data are collected and how they are used.

- Consistent with current HIPAA provisions, explicitly permit certain data collection practices and data uses that are essential to administering the health care system, providing the best possible care to the patient, and improving care for future patients.
- Provide opportunities for patients and consumers to assent to certain types of data collection and use, with clear guidelines regarding when and how such assent will be provided.
- Take into account that people's views regarding data collection and use depend on various factors, such as data type (sensitivity), the purpose of the original data collection, the uses to which the data will be put, and the characteristics of the organization(s) holding the data.

The goal of federal law should be to create guidelines sufficiently specific for all participants in the data ecosystem to rely on them while providing room for individuals to exercise their right to privacy.

3. All entities that hold health data should have clear policies

HIPAA creates clear policies regarding data collection, use, and sharing for certain entities. Despite its limitations, the clarity of the HIPAA approach should be emulated as we acknowledge the expanded health data environment. While basic parameters may exist in law, all organizations that hold health data should be required to have clear policies regarding how they will collect, use, and share health data, and these policies should be fully disclosed. The corollary to clear organizational policies is that any person providing health data should be able to readily determine which data are being collected, how they are being used, and if they are being shared. Entities holding health data must also be transparent about the revenue they obtain from data sale or reuse.



People should be presented with distinct choices regarding data sharing, and the implications of those choices should be presented as clearly as possible. When possible, goods and services should be available to people even if they choose not to share data.

Exceptions to honoring individual preferences must be tied to core health system functionality, as is currently the case with HIPAA, which provides exceptions for matters such as care coordination, public health surveillance, quality assessment and improvement, and administration.

Broad grants of authority to use, reuse, sell, and share data based on "check the box" processes should be prohibited.

4. Health sector leaders should advance a new covenant of health data use

Leading health systems, working with patients and their communities, should develop guidelines that reflect the mutual benefits to patients, clinicians, and communities that arise when health data are collected and used in accordance with ethical principles. Health systems should describe with clarity why they collect data and how that information will be used to advance patient and community health. Health systems should place explicit limitations on how and with whom the data will be shared so that patients can be confident their privacy will be respected.

This new covenant should reflect the reality that a learning health care system is only possible with substantial data use and data sharing. Patients can only obtain the best possible treatment if they are part of a learning system, and health systems can only provide the care patients expect if they have the data they need to continuously improve diagnosis and treatment.

Ethical members of the health care community must adopt limitations on data sharing that fall outside patient expectations, particularly with respect to selling data or sharing it with enterprises whose goals are unrelated to improving health. With those limitations in place, health care providers can reasonably expect patients to provide data for health- improving uses.

Health sector actors that collect, use, and share data should adopt this set of ethical principles and the specific guidelines they imply. The health sector should establish a mechanism of certification to allow the public to recognize systems that adopt these principles.

Health sector leaders should encourage all organizations that collect or hold health data—even those outside the traditional health care system—to adopt these principles.

These principles should form the basis for a robust education campaign directed at the public, patients, providers, and health system administrators regarding best practices in health data privacy.

5. Consumer participation in health data privacy practices should become the norm

All holders of health data should establish formal mechanisms for obtaining consumer input into their data policies and practices. The boundaries of legitimate action must not be set solely by those who possess health data or by law and regulation, but through processes that enable consumers and patients to assure that the value of privacy is considered and reflected in data use practices. With growing potential to make money through data collection, aggregation, and analysis, the test of whether health data should be used cannot solely be whether a business model will support it.

Patient and consumer engagement can come in various forms, such as through patient advisory boards and patient inclusion in organizational governance. Consumer engagement should emphasize participation by those most vulnerable to data misuse or exclusion from the benefits of data collection, use, and analysis.

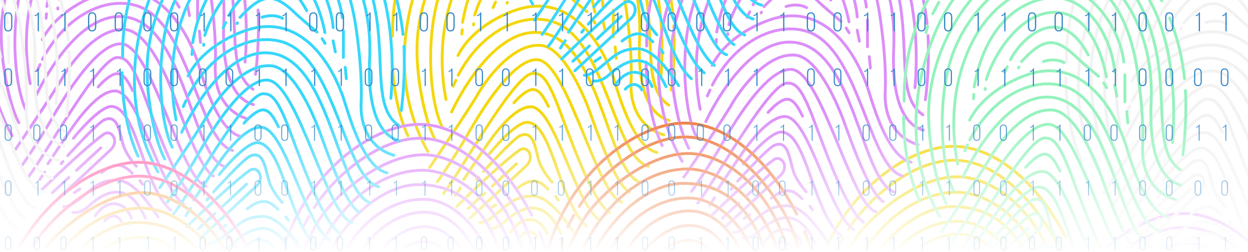
Moving Forward

Significant gains to human health are achievable if we harness the power of health data and rapidly improving analytics, yet current uses of health data go beyond what patients and consumers find acceptable. We must redesign our approach to health data privacy to honor the ethical value of privacy and to earn public support for using health data to positive ends.

The Aspen Health Strategy Group, with its multisector membership, has developed these ideas to motivate improvements in policy and practice. We call on Congress, the Biden administration, states, and the health sector to modernize the nation’s approach to health data privacy.

Reference

Institute of Medicine. (2007). The learning healthcare system: Workshop summary. National Academies Press. https://doi.org/10.17226/11903



BACKGROUND PAPERS

Health Information Privacy in the Digital Health Age

Deven McGraw, J.D., M.P.H., L.L.M.

Health Data Privacy in the Balance: Evolving Values and Priorities

Anita Allen, J.D., Ph.D.

Deidentification to Enhance Health Data Sharing

Bradley Malin, Ph.D.

The Value and Uses of Health Data in the Clinical Ecosystem

Kenneth D. Mandl, M.D., M.P.H.

Part 2

"To fully realize the potential of digital data and digital medicine, the United States needs comprehensive privacy and security protections, regardless of where the data are collected or maintained. At the same time, health protections must encourage and support responsible uses and disclosures."

– DEVEN MCGRAW, J.D., M.P.H., L.L.M.

Health Information Privacy in the Digital Health Age

Deven McGraw, J.D., M.P.H., L.L.M.

Introduction

Technological advances are changing the delivery of medicine and the pursuit of health and wellness. New digital technologies collect vast amounts of personal data from individuals in real time, both within and outside of traditional health care settings. This treasure trove of data includes information that looks like typical health data, such as a cancer diagnosis or blood glucose level, but also information that can be used to make health inferences and even predict whether an individual is likely to take medication as prescribed (Parker-Pope 2011). Health-relevant data can include information about social determinants of health, such as home ownership, job status, income, education levels, and access to nutritious food. It can also include internet search histories; genetic data from direct-to-consumer genetic testing companies; data collected by Fitbit, Apple Watch, and other wearable devices; and even information shared in Facebook social media groups or on Twitter (Figure 1). Because the relevance of a piece of data to describing a person's health depends more on how the data are used rather than the characteristics of the data themselves, it is difficult to know how much health-relevant data even exist.

Figure 1. Major Categories of Health-Relevant Data

CATEGORY	DEFINITION	EXAMPLES
Category 1	Generated by health care system	Electronic medical record data, prescriptions, laboratory data, including molecular "omics" data, pathology images, radiography, payer claims data.
Category 2	Generated by consumer health and wellness industry	Wearable fitness tracking devices, medical wearables such as insulin pumps and pacemakers, medical or health monitoring apps, patient-reported outcome surveys, direct-to-consumer tests (including DNA analysis) and treatments.
Category 3	Digital exhaust generated as a byproduct of consumer's daily activities	Social media posts, internet search histories, location, and proximity data.
Category 4	Non-health specific data: demographic, social, and economic sources	Race, gender, income, credit history, employment status, education level, residential ZIP code, housing status, census records, bankruptcy and other financial records, grocery store purchases, fitness club memberships, voter registration.

Source: McGraw & Mandl 2021

Much of the health-relevant data collected and shared in the United States is outside the scope of comprehensive privacy laws. This is particularly so for data collected through consumer applications (apps), which did not even exist at the time the Health Insurance Portability and Accountability Act (HIPAA), the nation's primary health data privacy law, was enacted. Numerous research reports have been published revealing how health apps routinely share data with third parties, with little transparency to users (Figure 2).

Figure 2. Examples of Privacy Research on Apps and Third-Party Data Sharing

SOURCE	DESCRIPTION
Kaldestad 2020	In a study of 10 apps (two of them intended to enable women to track menstrual cycles and predict ovulation times), researchers found the apps transmitted data on user activities in the app to 70 different third parties involved in advertising and profiling, without explicit consent from the users.
Test-Achats 2020	A study looked at 14 health and nutrition apps, including apps tracking medication use, migraines, and sleep and helping to manage diabetes, and found that all but one (Apple Health) shared data with third parties without full transparency to the user.
Huckvale et al. 2019	In a cross-sectional study, 29 of 36 apps for depression and smoking cessation transmitted data to services provided by Facebook or Google, but only 12 accurately disclosed this in a privacy policy.

Source: Author analysis

Along with concerns about increased collection and sharing of some health-relevant data, health and health care in the United States are also plagued by a lack of data to inform optimal clinical care, protect public health, drive medical discovery, and expand the evidence base for health and wellness interventions. Health outcomes in the United States trail those of peer industrialized countries across many domains (Papanicolas, Woskie, & Jha 2018). Adults receive recommended health care only a little over half the time (McGlynn et al. 2003), and much health care provided today is not supported by high-quality evidence (Califf et al. 2016). More robust collection and analysis of digital data are widely perceived to be vital to improving these outcomes and establishing what the National Academy of Medicine calls a “learning health care system” (Institute of Medicine Roundtable on Evidence-Based Medicine 2007).

Indeed, recent federal initiatives are pushing health care providers and health plans to share more health information, not less. The U.S. Department of Health and Human Services (HHS) issued a report in 2015 finding that health care providers and their vendors blocked information, declining to share it for essential purposes (HHS Office of the National Coordinator for Health IT 2015). Congress responded with provisions in the 21st Century Cures Act that established penalties for information blocking and directed HHS to take steps to assure the interoperability of health information. Federal agencies responded with initiatives for more widespread data sharing and the collection of health data. For example, the Centers for Medicare & Medicaid Services (CMS) now requires health plans under its purview to share claims data with subscribers and hospitals and to send alerts to physicians when their patients have been hospitalized. The Office of the National Coordinator for Health IT issued rules prohibiting information blocking and requiring certified electronic health records (EHRs) used by health care providers to adopt open, standard application programming interfaces to facilitate more seamless digital data sharing, including with individuals.

These initiatives focus on data sharing by entities within the traditional health care system. However, improving health and health care will also require using data generated by or within people's daily lives. There is increasing recognition that social determinants of health—factors outside of health care that affect people's lives (such as housing and food insecurity)—can have a large effect on health and wellness (Gottlieb, Sandel, & Adler 2013). Much of what influences an individual's health and well-being occurs outside the doctor's office or hospital (Quinn 2017), which means that the ability of individuals to collect and use health-relevant data to care for themselves and their loved ones is an essential part of a robust health data ecosystem.

The lack of strong, consistent protections for health data that respond to 21st-century risks could have the "long-term effect of reducing the uptake of new innovative technologies" and undermine the promise of digital medicine (Forbrukerrådet 2020). In this paper I argue that policy efforts to address health data have focused disproportionately on privacy—and within the realm of privacy, on consent and deidentification—and have failed to encourage beneficial uses of health data. I urge a multifaceted approach that incorporates protections and stimulates, encourages, or even demands responsible use as a mechanism to propel digital medicine initiatives in the United States.

Current Data Protections

The two main governing authorities related to health data privacy in the United States are HIPAA and the Federal Trade Commission (FTC) Act. While some states have privacy laws protecting health and personal data that are often more protective than federal law (Baum 2018), a discussion of state law protections is beyond the scope of this paper. HIPAA does not preempt more protective state laws.

Legal Framework

In 1996, Congress enacted the Health Insurance Portability and Accountability Act, intending to establish health insurance portability and reduce health care administrative costs by requiring submission of digital health care claims using standard formats. In the "Administrative Simplification" section of the law, Congress directed HHS to establish privacy and security protections for the digital health data that would be collected, used, and shared by health care providers and health plans as part of this digital claims submission process. In 2001, HHS promulgated these regulations, which are the source of most of HIPAA's protections. In 2009, as part of the Health Information Technology for Economic and Clinical Health Act (HITECH), Congress amended HIPAA to establish breach notification obligations and bolster protections for data already covered by HIPAA, in anticipation of the widespread adoption of electronic medical records by providers.

HIPAA
Health Insurance Portability and Accountability Act, a 1996 Federal law that restricts access to individuals private medical information
CONFIDENTIAL
FOR INTERNAL USE ONLY
PRESCRIPTION
℞

The Health Information Technology for Economic and Clinical Health Act
(HITECH Act)

The HIPAA privacy, security, and breach notification regulations provide a comprehensive set of protections, but only for some health data. HIPAA is a sectoral law; it covers only certain types of entities and generally does not extend to organizations and businesses outside the traditional health care ecosystem. (See Appendix for a brief summary of HIPAA's regulations.) In other words, HIPAA's coverage is mostly triggered by "who" (what entity is collecting, using, or sharing data) and far less by the type of data handled by these entities.

Thus, much of the health-relevant data collected from individuals, both actively and passively for a wide variety of purposes, resides outside of HIPAA's protections (Price & Cohen 2019; National Committee on Vital and Health Statistics 2019; U.S. Department of Health & Human Services 2016). For example, social media platforms, health and wellness apps, smartphones, life insurers, retailers, credit card companies, and internet search engines all collect health-relevant data outside the scope of HIPAA.

While HIPAA is the best-known health data privacy law, other federal laws protect health information in specific contexts. For example, federal rules known as Part 2 (a reference to their location in the Code of Federal Regulations) protect against the disclosure of identifiable information collected and used by federally supported substance use treatment programs. The Common Rule governs federally supported human subjects research, which includes research using identifiable personal information.

In addition to laws and regulations specifically focused on health, the primary authority in the United States when it comes to privacy is the Federal Trade Commission. The FTC exercises authority over privacy and security of personal information collected by businesses through its enforcement of Section 5(a) of the FTC Act, which broadly prohibits “unfair or deceptive acts or practices in or affecting commerce.”

The FTC Act applies to most businesses that collect data, including developers and marketers of mobile health technologies, social media sites, and internet search engines. Generally, however, the Section 5 authority in the law does not extend to nonprofit organizations or insurance companies, and there are some exceptions related to banks, savings and loan institutions, federal credit unions, and common carriers such as airlines.

The FTC’s broad authority extends to the collection, use, and disclosure of identifiable, personal data by covered businesses, including all health-relevant information described in Figure 1. The FTC has determined that its unfair and deceptive trade practices authority requires companies to honor their commitments set forth in privacy policies and terms of service and to be “fair” to consumers, including by adopting reasonable and appropriate data security practices (Solove & Hartzog 2011, p. 600).

Separately, the FTC administers the breach notification requirements of HITECH. Those requirements apply to “personal health records,” which are defined as an electronic record of identifiable health information that is “managed, shared, and controlled by or on behalf of the individual," and any applications that might be offered to users by that personal health record (for example, a nutrition management app).

Strengths of Existing Protections

HIPAA's comprehensive approach has significant strengths. The law creates enforceable boundaries for when and how identifiable information can be used and shared by the health care system. It does not place all of the obligations for protecting privacy on individuals deciding whether to provide consent. And, from its inception, HIPAA's regulatory framework has recognized that health data must be protected and made available for treatment, to secure payment, to enable health care institutions and medical practices to conduct operations, for public health and research purposes, and for patients to use for their own purposes. The FTC Act is comprehensive in its reach and sufficiently flexible to enable its application to a wide variety of types of businesses that collect personal data.

Weaknesses of Existing Protections

HIPAA's privacy and security rules were initially established in 2001; they have been amended minimally over the past two decades, except for amendments in 2013 arising out of HITECH. Although these regulations establish a comprehensive framework for protecting health data, it is not clear that they are sufficiently robust to meet the challenges of a 21st-century health data ecosystem (Butler 2017).

For example, HIPAA's rules for sharing data with contractors, and for sharing data that have been deidentified per HIPAA's standards, place few controls on what companies that lawfully receive data do with them. Entities covered by HIPAA, including vendor business associates, frequently sell data that are deidentified according to HIPAA standards but that can still be linked to create health profiles of individuals (Tanner 2017).

News reports about an arrangement between Google and Ascension Health to facilitate data analytics for Ascension caused an uproar, triggering investigations by HHS to assure that the arrangement complied with HIPAA (Garcia 2019). As a business associate to Ascension, Google therefore has the legal ability under HIPAA to use and share information from Ascension in ways similar to what health care providers and health plans can do, subject only to any contractual limitations Ascension may have imposed on Google.

Another example of how data are being used for commercial purposes is revealed in a class action lawsuit filed by a patient in June 2019 against the University of Chicago and Google. Here, the university sold supposedly deidentified medical record data to Google, enabling the company to create artificial intelligence tools that could be sold to physicians and hospitals (Cohen & Mello 2019).

These two examples focus on Google, but other large information technology companies—Facebook (Rohrer 2019), Microsoft (Thorne 2019), Amazon (Vena 2019), and Apple (CB Insights 2019)—have also announced initiatives involving the collection and use of health data, or indicated their intent to do so. While proponents of these arrangements cite the ability of these companies to bring their record of innovation in other sectors to "fix" health care (Watcher & Cassel 2020), their record of ubiquitous data collection and surveillance of consumers (Zuboff 2019), coupled with mishandling of personal information by Facebook (Davis 2019), Google (Nakashima 2018), and Twitter (Twitter Help Center n.d.), have generated some backlash.

In addition, although the FTC Act is applicable to most businesses collecting health-relevant data, many observers believe its protections are insufficient, in part because they depend too much on company commitments, and in part because they are not explicated in comprehensive regulations similar to the HIPAA rules (Terry 2020). For example, the FTC's recent settlement with Facebook regarding the company's failure to abide by a prior FTC consent decree and other alleged violations of the FTC Act has generated doubts about how seriously the FTC takes its enforcement role (Coldewey 2019). Research regarding the deficient privacy policies of mobile apps and reports about the mishandling of consumer data by tech companies reinforce concerns that the FTC and its authorities under the FTC Act are insufficient to address concerns about privacy for data outside HIPAA's boundaries.

Typical Federal Legislative Proposals and Their Limitations

Most recent legislation introduced in Congress to address the lack of privacy protections for personal data takes a sectoral approach, covering specific types of companies, while others are more comprehensive. For the most part, these bills would establish protections for personal data in general, not specifically for health data. Since personal data often are used for health purposes, these protections do have a direct impact on data collected in, for example, digital health tools.

In general, federal bills adopt one or more of the following approaches:

- requirements to provide individuals with clear notice about how their personal information is collected, used, and disclosed;
- requirements to provide individuals with choices (either opt-in or opt-out) for the collection, use, and disclosure of their personal information;
- broad definitions of personal data, with stricter standards for data considered to be deidentified (and therefore no longer regulated);
- establishment of individual rights concerning data, including the right to know whether a company possesses one's data, the right to request corrections, the right to obtain copies, and the right to have data deleted; and
- increased authority to, and resources for, the FTC to enforce new privacy mandates.

Opt-out

□

I do not allow my identifiable patient data to be shared outside of the GP practice for purposes except my own care.

OR

I do not allow the patient above's identifiable patient data to be shared outside of the GP practice for purposes exc
their own care

For the past five years, failure to reach compromise on two issues has stymied federal privacy legislation: whether the federal bill will preempt stronger state laws and whether the bill will allow individuals to bring private civil lawsuits to enforce the law (as opposed to the law's being enforced solely by governmental authorities) (Duball 2022). The U.S. Chamber of Commerce this year urged Congress to enact comprehensive privacy legislation establishing a single national standard (U.S. Chamber of Commerce 2022). However, preemption is unpopular with lawmakers who have concerns about replacing strong state privacy laws with a potentially weaker federal law, which is the most likely outcome in a sharply divided Congress. A private right of action is favored by lawmakers seeking to expand enforcement possibilities beyond governmental authorities; it is generally opposed by businesses.

The federal data privacy legislation generally proposed to date has important limitations, particularly as applied to health-relevant data.

Excessive Reliance on Notice and Consent

The predominant model for protecting privacy involves giving individuals notice of, and the right to consent to, uses and disclosures of their data. This model is widely recognized by privacy scholars as being inadequate to protect privacy (Hartzog & Richards 2018; Cate & Mayer-Schönberger 2013; Nissenbaum 2011; Pasquale 2014). Privacy notices are too long and hard to understand, and, in an age of big data, it is often difficult to predict all potential uses at the time of data collection. Particularly in online transactions, research has found that individuals often agree to terms of service without reading them (Berreby 2017). Companies design technology in ways that "maximize the collection, use, and disclosure of personal information," casting significant doubt on the idea that individuals truly can make informed choices online even when they are trying to do so (Hartzog 2018).

Reliance on notice and consent shifts the burden for protecting data privacy to the individual, rather than ensuring that institutions and data holders are accountable for acting in trustworthy ways. Companies can change their consent policies, and consumers may not be aware of these changes or feel they have little choice but to agree to them because they want to continue using a service. Finally, notice and consent do not protect against or remedy individual or even group harms, such as the use of data to discriminate against or devalue particular subpopulations.

The central data protection policy in the European Union, the Global Data Protection Regulation (GDPR), and consumer privacy legislation enacted in California, the California Consumer Privacy Act (CCPA), are strong privacy laws. Yet both rely significantly on consent (for GDPR, the right to consent before data are processed in most cases, and for CCPA, the right to opt out of data collection for purposes of targeted marketing). These laws do not appear to have significantly limited the ubiquitous collection and use of personal data by commercial enterprises.

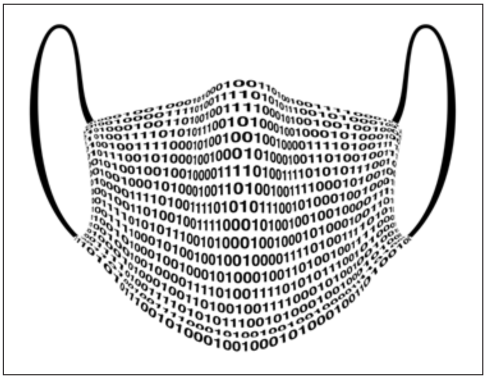
Overvaluing Deidentification

Most existing privacy laws cover only identifiable information. Information that has been deidentified, anonymized, or pseudonymized typically fall outside regulation. Although techniques to reduce identifiability of information lessen privacy risks, they do not reduce the risk to zero. For example, HIPAA's deidentification standard requires data to be at "very low" (not zero) risk of reidentification, yet HHS cannot hold recipients of deidentified data accountable for unauthorized reidentification (McGraw 2013). In addition to reidentification risk, some have raised concerns about the ethics of robust commercial sales of deidentified health data (Tanner 2017).

More recent privacy laws such as GDPR and CCPA appear to have more robust standards for how data qualify as deidentified or pseudonymized. For example, under the CCPA, data that can be linked to a particular person or household, such as through an IP address or advertising identifier, are covered by the law even if the individual is not identified. Because the CCPA is new, it is unclear whether these definitions will rein in commercialization of personal data. The CCPA was recently amended to address concerns that the stringent definition of identifiability would create obstacles to the use of health data in research (Kourinian, Nelson, & Martens 2020).

Failure to Support Data Availability

Responsible collection and analysis of health and health-relevant data are critical to addressing significant deficiencies in the U.S. health care system. However, most of the proposed health data privacy legislation focuses primarily on protecting personal data and not on promoting its availability. This may be of little importance for personal data in general, but it is a significant shortcoming when it comes to health-relevant data. As the COVID-19 pandemic has revealed, there is a need for data sharing in the health care system and for using data from nontraditional sources.



Surveys reveal that individuals practice "privacy-protective" behaviors, such as not seeking health care or hiding the truth about health conditions, if they do not trust that their information will be kept confidential (McGraw et al. 2009).

Thus, unless data are collected and used in ways that assure individuals their personal information will be handled responsibly and will not harm them, they are likely to opt out, formally or informally.

Why Not Just Extend HIPAA to Cover All Health Data?

A frequent suggestion is that extending HIPAA to protect all health-relevant data would be adequate. While elegant in its simplicity, this suggestion would do little to address the risks associated with the use of data by companies outside the traditional health care system. HIPAA's protections and regulations were deliberately crafted to accommodate the ways that health care providers and health plans need to access, use, and disclose health information in order to operate. For example, HIPAA includes a long list of permitted uses and disclosures to support core health care activities such as treatment, payment, and operations (to name just a few). Those provisions would not effectively govern companies outside health care, many of which do not treat, pay, or perform other functions characteristic of health care providers or health plans.

Also, a critical privacy concern is the ubiquitous collection of personal data by businesses, but HIPAA regulations include no restrictions on what information an entity covered by HIPAA can collect. And there are no outright prohibitions on what entities covered by HIPAA can do with data, as uses or disclosures that are not expressly permitted can still occur with the written authorization of the individual who is the subject of the data. Simply extending the reach of HIPAA would do nothing to address these limitations in the law's ability to protect health data privacy.

Building a Health Data Ecosystem through Strong Federal Policies

The dual needs in health care both to protect data and assure their availability call for comprehensive policies governing all entities collecting and using health and health-relevant information, whether covered by HIPAA or not. HIPAA's provisions are not a perfect fit for settings outside the traditional health care ecosystem, but its approach, which relies less on consent and more on setting expectations for data holders, is a worthy model. Congressional action is needed, and policymakers should consider the following in crafting comprehensive policies to govern health and health-relevant information.

Establish Rules Based on Reasonable Consumer Expectations

Instead of relying on consent, policymakers should establish some limits on the collection, use, and disclosure of health information by businesses. For example, the FTC has recommended that "companies should limit data collection to that which is consistent with the context of a particular transaction or the consumer's relationship with the business, or as required or specifically authorized by law" (Federal Trade Commission 2012). In other words, data collection should be limited to what a consumer might expect, given the context.

HIPAA's regulations contain few limits on how entities collect health information, choosing instead to comprehensively regulate how that information can be used and disclosed once it is in the hands of an entity covered by HIPAA. This is unusual for privacy law. HIPAA is based on "fair information practice principles," which are the foundation for U.S. and international information privacy laws, and collection limitations are a vital part of those principles (Gellman 2022). When HHS first drafted the HIPAA regulations, it may have made sense to disregard collection limitations, as

HHS was setting ground rules for how a defined set of entities within the health care system handle data. However, for commercial enterprises, some limits on the collection of health and health-relevant data may make sense. For example, the collection of health and health-relevant data should be prohibited unless the data collection is consistent with consumer expectations and intended to benefit the individual or population health.

Use and disclosures of health-relevant data similarly should be limited to what the consumer would reasonably expect, given the context. This maxim should also govern the repurposing of information. For example, technology and telecommunications companies routinely collect geolocation data; governments around the world are seeking or are already collecting these data for COVID-19 response activities. These data were not collected initially for this purpose, and consumers likely did not expect their data to be used that way. At a minimum, companies should obtain clear consent for potentially beneficial (or at least nonharmful) data uses that go beyond what the consumer would reasonably expect given the context.

Consumer Oversight

Companies should establish independent data ethics review boards. Such boards would evaluate the legal and ethical implications of proposed data projects, as well as their potential to improve health or the health care system (Parasidis, Pike, & McGraw 2019). Such boards are similar to Institutional Review Boards (IRBs), which provide an independent review of proposals for research on human subjects under federal law. However, the data ethics review boards would focus more on privacy than the potential for physical or mental harm from research and include members with substantial privacy expertise. Further, data ethics review boards would evaluate uses and disclosures beyond those for research.

For such boards to be effective they should be properly constituted and independent, with members drawn from outside the company, including consumers, patients, or both. They should also have direct reporting channels to governing boards, such as a company’s board of directors. Facebook recently announced the establishment of an independent oversight board to achieve “fair decision-making” concerning the removal of unacceptable content on the site. Among the board’s authorities are to “instruct” Facebook to allow or remove content and “interpret” Facebook’s community standards and other policies “in light of Facebook’s articulated values” (Facebook 2019).

Formal Assessment

GDPR requires a data protection impact assessment, and in some cases regulatory review, for certain types of data processing, particularly methods using new technologies that are "likely to result in a high risk to the rights and freedoms of natural persons" (GDPR Article 35 2022). Similarly, federal agencies in the United States are required to conduct privacy impact assessments "for all new or substantially changed technology" that collects, maintains, or disseminates personally identifying information (U.S. National Archives and Records Administration 2022). Such assessments could have value if they are periodically subject to independent, objective review and not merely check-the-box exercises.

Data Trusts

Some have proposed data trusts to assure that companies use and disclose personal data for the benefit of consumers. The term "data trust" does not have a universally accepted meaning but generally refers to mechanisms for collecting and sharing data that are characterized by binding data governance rules enforced as a matter of trust law. Proponents see data trusts or civic trusts as mechanisms for assuring that companies continue to honor their data commitments to consumers regardless of changes in company strategy or sale of the company (McDonald & Porcaro 2015). Consumer data trusts have also been defined as "intermediaries that aggregate consumers' interests and represent them vis-à-vis data-using organizations" (Stiftung Neue Verantwortung 2020). By aggregating consumer interests, consumer trusts would have the bargaining power to negotiate better terms for data use and disclosure than could be achieved by any individual consumer. Existing laws giving individuals the right to copies of their information (for example, HIPAA and GDPR) could facilitate the establishment of these trusts, as individuals could direct these trusts to hold and manage their information. But even proponents of trusts counsel that the protective value of trusts depends upon who sets the data governance rules.

Increased Oversight

Other options include requiring companies that collect or process health or health-relevant data to adhere to additional oversight and other requirements. Ontario, Canada, permits "data custodians" (those who hold the data) to disclose personal health information for the purpose of health system improvement, but only to entities

approved by a privacy commissioner to have practices and procedures in place that protect privacy and maintain confidentiality (Information and Privacy Commissioner/Ontario 2004). Building on that theme, health data collection and processing could be limited only to entities that demonstrate through periodic audits that they meet ethical, privacy, and security standards. Companies collecting health and health-relevant data also could be required to segment or firewall their health businesses from other aspects of the company.

Increased Transparency

Although notice and consent should not be the cornerstone of privacy, individuals still want and expect to have notice of, and some choice about, collection, use, and disclosure of health and health-relevant information (Ogury 2019). The FTC has recommended "simplified choice," with clearer, shorter, and more standardized privacy notices in circumstances where the data collection, use, and sharing are beyond what consumers would ordinarily expect or where sensitive data are involved (Federal Trade Commission 2012). For example, companies can improve notice and choice through layered notice and the use of visuals to improve comprehension (Schaub et al. 2015; Kay & Terry 2010).

Even if consent is not sought for a particular use or disclosure, either because it is within consumer expectations or is mandated or authorized by law, companies should still be required to be fully transparent about data uses and disclosures (ABA Banking Journal 2019). As demonstrated by the uproar over Google's arrangements with the University of Chicago Medical Center and Ascension Health System, the black box nature of health data uses and disclosures, including sales of deidentified data, has the potential to deepen consumer mistrust of digital medicine technologies.

Strengthen Remedies for Privacy Harms

In the past, the FTC has rejected calls for a harm-based model of privacy that focuses only on protecting consumers from harms such as "physical security, economic injury, and unwanted intrusions into their daily lives" (Federal Trade Commission 2012). The FTC concluded that such a model would fail to recognize "a wider range of privacy-related concerns, including reputational harm or the fear of being monitored."

Nevertheless, harm should be considered when addressing privacy concerns. Feelings of risk and anxiety are among the harms suffered by individuals whose data are breached (Solove & Citron 2018). Rules-based privacy regimes like HIPAA create enforceable expectations about how health data must be handled, regardless of whether individuals suffer any cognizable harm when organizations don't follow the rules. In enforcing HIPAA, HHS considers whether a violation harmed individuals when determining the level of civil monetary penalty it will pursue. In HITECH, Congress amended the HIPAA Privacy Rule to require HHS to establish a mechanism for individuals who have been harmed by HIPAA violations that HHS chooses to pursue to receive a portion of any resulting civil monetary penalties or settlements. However, HHS has yet to act on this measure.

One interesting example of a harm-based privacy measure is a privacy tax on data collectors and processors that could fund no-fault compensation for privacy harms (Edwards 2004). Companies could be required to establish compensatory funds that broadly recognize the harms that can occur both to individuals and groups as a result of unauthorized or unethical uses or disclosures of data.

Maintain Regulation of Deidentified Data

Regulation of health-relevant data should provide incentives for the use and disclosure of data in less identifiable forms. However, given that these data retain some residual risk of reidentification, they should also be subject to some regulation. For example, civil monetary penalties should be imposed for unauthorized reidentification of deidentified data, and criminal penalties imposed for intentional reidentification. Relaxing, but not eliminating, regulations on data at very low risk of reidentification provides incentives for entities to collect, use, and disclose deidentified data.

But merely controlling for the risk of reidentification will not be sufficient to garner consumer trust in how companies handle their health and health-relevant data.

Companies should be required to be transparent about the uses and disclosures of deidentified data, identify the general methods used for deidentification, and provide consumers with some choices about disclosures. Disclosures of deidentified data could also be subject to an ethics board review.

An Interim Solution: Adoption of Enforceable Best Practices

Even in the absence of federal legislation, companies collecting health-relevant data should adopt best practices, with commitments enforceable by the FTC when the company is covered by the FTC Act. Best-practice frameworks have recently been published by a joint effort of the Center for Democracy & Technology and the Executives for Health Innovation (eHI), and by the American Medical Association, the CARIN Alliance, and the Consumer Technology Association. As well, many of the policy recommendations noted in this paper could be adopted as best practices.

Another best practice of health systems today is data use agreements that bind data recipients to contractual commitments (McGraw & Mandl 2021). These agreements often include prohibitions on further use and disclosure, and, in the case of deidentified data, commitments not to reidentify; HIPAA only requires the use of such agreements in limited circumstances. The success of such an approach depends on the parties agreeing to responsible terms, which does not always happen when one party has greater bargaining power. Further, only parties to a contract can enforce the contract's terms. While such contracts can be protective, they also can be vehicles for protecting data as a proprietary asset, which can limit their availability, even for potentially beneficial uses.

Conclusion

To fully realize the potential of digital data and digital medicine, the United States needs comprehensive privacy and security protections, regardless of where the data are collected or maintained. At the same time, health protections must encourage and support responsible uses and disclosures. Privacy legislation tends to focus more on protecting health and health-relevant data than on assuring its appropriate use. Proposed measures for protecting data rely too much on notice and consent and on deidentification. What is needed instead is a multipronged approach that leverages fair information practice principles to implement strong privacy protections but also includes accountability, even for the use of deidentified or anonymized data. Such measures should also assure the availability of health and health-relevant data for societal benefit. Even before such legislation is enacted, companies should adopt best practices, with enforcement of those practices by the FTC.

Deven McGraw, J.D., M.P.H., L.L.M. is the lead for Data Stewardship and Data Sharing at Invitae, the company that acquired Ciitizen, the platform she founded for patients to gather their health information. From 2015 to 2017, she directed U.S. health privacy and security as deputy director of health information privacy at the HHS Office for Civil Rights and as chief privacy officer (acting) of the Office of the National Coordinator for Health IT. Widely recognized for her expertise in health privacy, McGraw directed the Health Privacy Project at the Center for Democracy & Technology for six years and led the privacy and security policy work for the HITECH Health IT Policy Committee. She also served as the chief operating officer of the National Partnership for Women and Families and advised health industry clients on HIPAA compliance and data governance while a partner at Manatt, Phelps & Phillips, LLP.

Appendix: Overview of HIPAA's Regulatory Provisions

HIPAA governs a wide range of identifiable "protected health information" (PHI), which is broadly defined and includes demographic and other information related to current or past health status that is created, held, or transmitted by an entity covered by HIPAA. However, in terms of "who" is covered by the law, HIPAA's scope is narrow. In essence, it applies to most health care providers (except those that do not engage in payment-related transactions using HIPAA standards [https://www.hhs.gov/hipaa/for-professionals/covered-entities/index.html]), health plans and health care clearinghouses, and their contractors (known as "business associates"). When HIPAA applies, the Privacy Rule includes detailed provisions regarding how PHI, in digital, paper, or other forms, can be used and disclosed, such as for treatment, payment, public health, and research, and specifies when an entity needs to obtain the prior authorization of the data subject. These permitted uses and disclosures–allowed without the need to obtain consent or authorization of the data subject–are largely designed to accommodate data flows within the traditional health care ecosystem. The list of such uses and disclosures includes:

- Treatment (45 CFR §164.502(a)(1)(ii) and §164.506(a))
- Payment and payment-related activities (45 CFR §164.502(a)(1)(ii) & §164.506(a))
- Health care operations (45 CFR §164.500, §164.502(a)(1)(ii), and §164.506(a))
- Quality assessment and improvement activities
- Population-based activities relating to improving health or reducing costs
- Case management and care coordination
- Reviewing the competence of health care professionals, evaluating provider performance, training of health care professionals, and licensure/certification/accreditation activities
- Underwriting and other activities related to health insurance
- Medical review, legal, and auditing, including fraud and abuse detection and compliance
- Business planning and development
- Business management and general administrative activities (including fundraising for the benefit of the covered entity and sale or transfer of covered entity assets)
- To public health authorities for public health purposes (45 CFR §164.512(b))
- To business associates (and sub-business associates [provided a HIPAA-compliant business associate agreement {BAA} is executed] (45 CFR §164.502(e)(1))
- Where required by other law (such as a state law mandating disclosure of health information) (45 CFR §154.512(a)): health care oversight (to health oversight agencies) (45 CFR §164.512(d))
- To avert a serious threat to health and safety (45 CFR §164.512(j))
- As part of judicial and administrative proceedings (45 CFR §164.512(e))
- For disaster relief (to disaster relief organizations) (45 CFR §164.510(b)(4))
- Law enforcement (subject to conditions) (45 CFR §164.512(f))
- For national security (45 CFR §164.512(k)(2))
- Disclosures about victims of abuse and neglect (45 CFR §164.512(c))
- For tissue or organ donation purposes (45 CFR §164.512(h))
- To coroners, medical examiners, funeral directors (45 CFR §164.512(g))
- For research, if the need for the data subject's authorization is waived by an Institutional Review Board or Privacy Board (45 CFR §164.512(i))

HIPAA's protections do not follow the data: once they are disclosed outside of an entity covered by HIPAA, they only continue to be covered by HIPAA if they are received or collected by an entity also subject to HIPAA.

HIPAA also establishes rights for individuals, including the right to obtain a copy of PHI and to request amendments to these data. The Security Rule establishes baseline physical, technical, and administrative safeguards that apply to electronic PHI, and the Breach Notification Rule requires notification of individuals and regulators in the event of breaches of PHI. HIPAA also defines deidentified data, and sets standards on how to achieve it, but places no limits on its use or disclosure, regardless of who controls the information.

References

ABA Banking Journal. (2019, June 11). Study: Consumers increasingly concerned with data security, privacy. https://bankingjournal.aba.com/2019/06/study-consumers-increasingly-concerned-with-data-security-privacy/

Baum, S. (2018, November 13). Navigating state patient data privacy laws will only get more challenging. MedCity News. https://medcitynews.com/2018/11/navigating-state-patient-data-privacy-laws-will-only-get-more-challenging/

Berreby, D. (2017, March 3). Click to agree with what? No one reads terms of service, studies confirm. *The Guardian*. https://www.theguardian.com/technology/2017/mar/03/terms-of-service-online-contracts-fine-print

Butler, M. (2017, April). Is HIPAA outdated? While coverage gaps and growing breaches raise industry concern, others argue HIPAA is still effective. *Journal of AHIMA,* 88(4), 14–17, 52. https://bok.ahima.org/doc?oid=302073#.Y1b6j3bMK5c

Califf, R. M., Robb, M. A., Bindman, A. B., Briggs, J. P., Collins F. S., Conway, P. H., Coster, T. S., Cunningham, F. E., De Lew, N., DeSalvo, K. B., Dymek, C., Dzau, V. J., Fleurence, R. L., Frank, R. G., Gaziano, J. M., Kaufmann, P., Lauer, M., Marks, P. W., McGinnis, J. M., . . . Sherman, R. E. (2016, December 15). Transforming evidence generation to support health and health care decisions. *New England Journal of Medicine*, 375, 2395–2400. doi: 10.1056/NEJMsb1610128

Cate, F. H., & Mayer-Schönberger, V. (2013). Notice and consent in a world of big data. Indiana University Bloomington, Maurer School of Law Digital Repository. https://www.repository.law.indiana.edu/facpub/2662

CB Insights. (2019, January 8). Apple is going after the healthcare industry, starting with personal health data. https://www.cbinsights.com/research/apple-healthcare-strategy-apps/

Cohen, I. G., & Mello, M. M. (2019, September 24). Big data, big tech, and protecting patient privacy. *JAMA*, 322(12), 1141–1142. doi: 10.1001/jama.2019.11365

Coldewey, D. (2019, July 24). 9 reasons the Facebook FTC settlement is a joke. TechCrunch. https://techcrunch.com/2019/07/24/9-reasons-the-facebook-ftc-settlement-is-a-joke/

Congressional Research Service. (2019). Data protection law: An overview. https://fas.org/sgp/crs/misc/R45631.pdf

Davis, J. (2019, February 20). Facebook accused of exposing user health data in complaint to FTC. *Health IT News*. https://healthitsecurity.com/news/facebook-accused-of-exposing-user-health-data-in-ftc-complaint

Duball, J. (2022, June 6). US lawmakers unveil bipartisan American Data Privacy and Protection Act. *Privacy Advisor.* https://iapp.org/news/a/congress-unveils-american-data-privacy-and-protection-act/

Edwards, L. (2004). Reconstructing consumer privacy online: A modest proposal. *International Review of Law, Computers and Technology, 18*, 313–334. doi:10.1080/1360086042000276762

Facebook. (2019). Facebook Oversight Board charter, section 4. https://about.fb.com/wp-content/uploads/2019/09/oversight_board_charter.pdf

Federal Trade Commission. (2012). Protecting consumer privacy in an era of rapid change: Recommendations for businesses and policymakers. https://www.ftc.gov/reports/protecting-consumer-privacy-era-rapid-change-recommendations-businesses-policymakers

Forbrukerrådet. (2020). Out of control: How consumers are exploited by the online advertising industry. https://www.conpolicy.de/en/news-detail/out-of-control-how-consumers-are-exploited-by-the-online-advertising-industry/

Garcia, A. (2019, November 12). Google's "Project Nightingale" center of federal inquiry. CNN. https://www.cnn.com/2019/11/12/tech/google-project-nightingale-federal-inquiry/index.html

GDPR (Global Data Protection Regulation) Article 35. (2022). https://gdpr.eu/data-protection-impact-assessment-template/

Gellman, R. (2022). Fair information practices: A basic history. bobgellman.com. https://bobgellman.com/rg-docs/rg-FIPshistory.pdf

Gottlieb, L., Sandel, M., & Adler, N. E. (2013, June 10). Collecting and applying data on social determinants of health in health care settings. *JAMA Internal Medicine,* 173(11), 1017–1020. doi: 10.1001/jamainternmed.2013.560

Hartzog, W. (2018). *Privacy's blueprint: The battle to control the design of new technologies.* Harvard University Press.

Hartzog, W., & Richards, N. (2018, December 20). It's time to try something different on internet privacy. *Washington Post.* https://www.washingtonpost.com/opinions/its-time-to-try-something-different-on-internet-privacy/2018/12/20/bc1d71c0-0315-11e9-9122-82e98f91ee6f_story.html?noredirect=on

HHS Office of the National Coordinator for Health IT. (2015). Report to Congress: Report on information blocking. https://www.healthit.gov/sites/default/files/reports/info_blocking_040915.pdf

Huckvale, K., Torous, J., & Larsen, M. E. (2019, April 19). Assessment of the data sharing and privacy practices of smartphone apps for depression and smoking cessation. *JAMA Network Open,* 2(4), e192542. doi:10.1001/jamanetworkopen.2019.2542

Information and Privacy Commissioner/Ontario. (2004). A guide to Personal Health Information Protection Act. https://www.ipc.on.ca/wp-content/uploads/Resources/hguide-e.pdf

Institute of Medicine (U.S.) Roundtable on Evidence-Based Medicine. (2007). L. Olsen, D. Aisner, and J. McGinnis (Eds.). *The learning healthcare system: Workshop summary.* National Academies Press.

Kay, M., & Terry, M. (2010). Textured agreements: Re-envisioning electronic consent. Symposium on Usable Privacy and Security (SOUPS) 2010, Redmond, WA. https://www.scholars.northwestern.edu/en/publications/textured-agreements-re-envisioning-electronic-consent

Kourinian, A., Nelson, P., & Martens, K. (2020, November 5). Amendment to CCPA harmonizes privacy and healthcare information requirements—exemptions for de-identified patient information under AB 713 address HIPAA and CCPA standards. JDSupra. https://www.jdsupra.com/legalnews/amendment-to-ccpa-harmonizes-data-72076/

McDonald, S., & Porcaro, K. (2015, August 4). The civic trust. Medium. https://medium.com/@digitalpublic/the-civic-trust-e674f9aeab43

McGlynn, E. A., Asch, S. M., Adams, J., Keesey, J., Hicks, J., DeCristofaro, A., & Kerr, E. A. (2003, June 26). The quality of health care delivered to adults in the United States. *New England Journal of Medicine,* 348, 2635–2645. doi: 10.1056/NEJMsa022615

McGraw, D. (2013, January–February). Building public trust in uses of Health Insurance Portability and Accountability Act de-Identified data. *Journal of the American Medical Informatics Association, 20*(1), 29–34. doi: 10.1136/amiajnl-2012-000936

McGraw, D., Dempsey, J. X., Harris, L., & Goldman, J. (2009). Privacy as enabler, not an impediment: Building trust into healthcare information exchange. *Health Affairs, 28*(2), 416–427.

McGraw, D., & Mandl, K. D. (2021). Privacy protections to encourage use of health-relevant digital data in a learning health system. npj Digital Medicine, 4(2). https://doi.org/10.1038/s41746-020-00362-8

McGraw, D., & Petersen, C. (2020, March). From commercialization to accountability: Responsible health data collection, use, and disclosure for the 21st century. *Applied Clinical Informatics, 11*(2), 366–373. doi: 10.1055/s-0040-1710392

Nakashima, R. (2018, August 13). AP exclusive: Google tracks your movements, like it or not. Associated Press. https://apnews.com/828aefab64d4411bac257a07c1af0ecb

National Committee on Vital and Health Statistics (NCVHS). (2019). Health information privacy beyond HIPAA: A framework for use and protection (a report for policy makers). U.S. Department of Health and Human Services. https://ncvhs.hhs.gov/wp-content/uploads/2019/07/Report-Framework-for-Health-Information-Privacy.pdf

Nissenbaum, H. (2011, Fall). A contextual approach to privacy online. *Daedalus*, 140(4), 32–48. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2567042

Ogury. (2019, October 28). How consumers really feel about their privacy and data. https://ogury.com/blog/how-consumers-really-feel-about-their-privacy-and-data/

Papanicolas, I., Woskie, L. R., & Jha, A. K. (2018, March 13). Health care spending in the United States and other high-income countries. *JAMA, 319*(10), 1024–1039. doi:10.1001/jama.2018.1150

Parasidis, E., Pike, E., & McGraw, D. (2019, April 18). A Belmont report for health data. *New England Journal of Medicine*, 380(16), 1493–1495. doi: 10.1056/NEJMp1816373.

Parker-Pope, T. (2011, June 20). Keeping score on how you take your medicine. *New York Times.* https://well.blogs.nytimes.com/2011/06/20/keeping-score-on-how-you-take-your-medicine/

Pasquale, F. (2014). Redescribing health policy: The importance of information policy. *Houston Journal of Health and Law Policy,* 14, 95–128. https://www.law.uh.edu/hjhlp/volumes/Vol_14/Pasquale.pdf

Price, W. N., II, & Cohen, I. G. (2019, January 7). Privacy in the age of medical big data. *Nature Medicine,* 25, 37–43. https://www.nature.com/articles/s41591-018-0272-7

Quinn, M. (2017, February 17). The future of health care is outside the doctor's office. Governing. https://www.governing.com/topics/health-human-services/gov-community-health-workers.html

Rohrer, V. (2019, November 4). Facebook is getting into healthcare: Is that a good idea? Motley Fool. https://www.fool.com/investing/2019/11/04/facebook-is-getting-into-healthcare-is-that-a-good.aspx

Schaub, F., Balebako, R., Durity, A. L., & Cranor, L. F. (2015). A design space for effective privacy notices. 2015 Symposium on Usable Privacy and Security. https://www.usenix.org/system/files/conference/soups2015/soups15-paper-schaub.pdf

Solove, D., & Citron, D. K. (2018). Risk and anxiety: A theory of data-breach harms. *Texas Law Review,* 96, 737–786. https://scholarship.law.bu.edu/faculty_scholarship/616

Solove, D. J., & Hartzog, W. (2011). The FTC and the new common law of privacy. *Columbia Law Review,* 114, 583–676.

Stiftung Neue Verantwortung. (2020). Designing data trusts. Why we need to test consumer data trusts (policy brief). https://www.stiftung-nv.de/en/publication/designing-data-trusts-why-we-need-test-consumer-data-trusts-now

Tanner, A. (2017). *Our bodies, our data: How companies make billions selling our medical records.* Beacon Press.

Terry, N. P. (2020, March). Assessing the thin regulation of consumer-facing health technologies. *Journal of Law and Medical Ethics,* 48(1 suppl.), 94–102. doi: 10.1177/1073110520917034.

Test-Achats. (2020). Nutrition and health applications do not respect privacy. https://www.test-achats.be/hightech/internet/presse/les-applications-nutrition-sante-ne-respectent-pas-la-vie-privee

Thorne, J. (2019, February 7). Microsoft Healthcare reveals more of its strategy with new cloud and AI products for hospitals. GeekWire. https://www.geekwire.com/2019/microsoft-healthcare-gets-ready-prime-time-cloud-ai-products/

Twitter Help Center. (n.d.). Personal information and ads on Twitter. Retrieved on November 2, 2022 from https://help.twitter.com/en/information-and-ads#10-08-2019

U.S. Chamber of Commerce. (2022, January 13). Letter sent to members of the U.S. Congress on national privacy legislation. https://www.uschamber.com/technology/data-privacy/coalition-letter-on-national-privacy-legislation

U.S. Department of Health & Human Services. (2016). Examining oversight of the privacy and security of health data collected by entities not regulated by HIPAA. https://www.healthit.gov/sites/default/files/non-covered_entities_report_june_17_2016.pdf

U.S. National Archives and Records Administration. (2022). Privacy impact assessments. https://www.archives.gov/privacy/privacy-impact-assessments

Vena, D. (2019, December 14). Amazon continues to make stealth moves into health care. Motley Fool. https://www.fool.com/investing/2019/12/14/amazon-continues-to-make-stealth-moves-into-health.aspx

Wachter, R. M., & Cassel, C. K. (2020, February 11). Sharing health data with digital giants: Overcoming obstacles and reaping benefits while protecting patients. *JAMA*, *323*(6), 507–508. doi: 10.1001/jama.2019.21215

Zuboff, S. (2019). *Surveillance capitalism: The fight for a human future at the new frontier of power.* PublicAffairs.

“Notwithstanding the inequitable distribution of privacy protections, and the fact that conditions of privacy can be exploited to conceal serious wrongdoing by institutions and individuals, I believe health data privacy merits the robust protection of law (Allen 2011).”

– ANITA L. ALLEN, J.D., PH.D.

Health Data Privacy in the Balance: Evolving Values and Priorities

Anita L. Allen, J.D., Ph.D.

Introduction

Attitudes toward health data privacy appear to be rapidly changing in the United States. Openness and sharing are welcomed, replacing secrecy and confidentiality as the norm. Ordinary people increasingly share personal health information, whether in face-to-face conversations, on social media, through digital health apps, or by accepting cookies as they browse medical information online (Friedman et al. 2022; Allen 2016). Researchers, businesses, and government officials are keen to exploit available health-relevant data and encourage sharing in the interest of consumer education, product marketing, health research, improved clinical care, and more (Denny et al. 2019). Excitement surrounds the push to apply machine learning and data analytics to health and medicine, from creating a "learning health care system" to the study of behavior on social media to aid diagnoses of depression (McGraw & Mandl 2021; Horvitz & Mulligan 2015). Artificial intelligence whose data inputs and methods are not visible to users (black-box analytics) potentially recommend superior approaches to our care, make decisions about what we can afford, and assess our worth for credit and insurance (Ford & Price 2016; Pasquale 2015; Waldman 2021). Traditional health privacy values can seem out of step when innovative, data-driven technology promises improved knowledge and efficiency (Romoser 2018; Harris 2021).

Data privacy and data disclosure are sometimes presented as horns of an ethical policy dilemma (Black 2003). Calls for policy reforms that balance privacy and disclosure, rather than prioritizing either, are at least superficially attractive methods for resolving the apparent conflict (McGraw & Mandl 2021). Indeed, balancing as the appropriate response to the dilemma is only rarely questioned (Bayer & Fairchild 2010). While seeking balance can be a useful deliberative technique, in the present complex policy arena we must not allow talk of balancing to cloak the depth of lost faith in a gospel of privacy that is still wise, even though it has become unfashionable (Deapen 2006).

Much is at stake in the shift from health data privacy to health data disclosure as a dominant norm and preference. In this paper I describe the shift, explore reasons for it, and consider some related ethical and legal implications. I also note the significance of the shift as it relates to the quest for resolving health disparities and systemic injustices (McGraw & Mandl 2021). Observing the move from emphatic privacy to emphatic disclosure—a profound change in public policy priorities—bioethicists are called upon to rethink the bases for regarding strong protection of health information privacy as a paramount policy aspiration and imperative.

The Case for Health Data Privacy

Health data remain among the most sensitive category of personal information, along with financial and educational information. Although the disclosure of health information can be beneficial, and information sharing has become more common in recent decades, individuals keep some health matters to themselves. When we share sensitive health concerns, most individuals employ culturally appropriate intimacy and discretion (Allen 2011; 2021b). We are particular about who knows what because we want the social power and other advantages that health information privacy confers. Our families, friends, employers, employees, coworkers, doctors, researchers, and governments may not acquire all of the health information about us that they desire (Allen 2003; 2007).

There are many reasons to take health data privacy seriously. Confidentiality in the provider-patient relationship likely encourages people to seek medical attention and discuss their symptoms and behaviors frankly. In the wrong hands, health data can lead to shame, embarrassment, stigma, censure, discrimination, and interference with autonomous decision-making (Allen 2007; 2014; 2021b).

Large health industry data breaches by nefarious actors increase the risk that "unscrupulous individuals may use health information to prey upon others through a variety of methods, from marketing ineffective remedies or fraudulent financial opportunities, to burglarizing unoccupied homes of persons who are hospitalized" (Deapen 2006, p. 634). There are cultural concerns as well. Health data may be used by well-meaning researchers or health care providers for purposes contrary to the beliefs, values, and interests of a tribe or racial group, as the 2004 Havasupai Indian DNA use case against the Arizona Board of Regents revealed (Garrison 2013; Obermeyer et al. 2019). Without adequate health data privacy, medical patients, wellness product consumers, and research subjects from all backgrounds and demographics are at the mercy of those who would exploit and misuse their data.

The United States possesses a vast array of state and federal laws protecting physical, informational, decisional, associational, intellectual, and proprietary dimensions of privacy, a portion of which relates to health (Allen & Rotenberg 2016; Allen 1997). Much of this law predates the digital economy. It is based on the notion that privacy is a kind of freedom or liberty, achieved in the name of dignity or respect for persons, by granting individuals control over their personal data (Allen 2011).

Yet as a valuable, legally protected resource and source of social power, the ability to experience privacy is inequitably distributed (Skinner-Thompson 2020; Véliz 2020). The practical ability of individuals to take advantage of legal privacy protections varies with their level of education, knowledge of data management practices, and access to digital technology. Race, income, gender, and sexual orientation have a role in determining how much privacy a person can expect in everyday life and how much they can benefit from the privacy laws that in theory pertain to everyone equally (Allen 2022). Data protection laws, which are products of constitutions, statutes, and common law, have social dimensions that I call "compliance limitations" (Allen 2015). Individuals and institutions required by law to protect data by concealment and nondisclosure may lapse when it comes to the information of certain culturally salient groups. Uneven obedience to privacy law may expose individuals in disadvantaged social categories—including racial minority groups, low-income groups, and women—to harm, constituting a compliance limitation of the law.

For example, African Americans have the same privacy rights on paper as other racial groups but are especially vulnerable to diminished privacy in the form of oversurveillance by police (Allen 2022; Arnett 2020). People of all races with low levels of income who depend on government services and benefits are burdened by more accountability requirements and enjoy less informational privacy than affluent people who are securely housed, employed, and fed (Bridges 2017; Allen 1988). Adults who are in active military service, judged mentally incompetent, or incarcerated, as well as minors, lack substantial control over their privacy. Women are more vulnerable to privacy abuses than men, due to a subordinate social standing that invites noncompliance with ethical rules and laws governing privacy invasion and battery (Allen 2000; Citron 2014). This can be seen in the health sector where female patients have been physically abused by their physicians purporting to perform gynecological exams (Allen 2015).

Governmentally implemented privacy resources may not serve all groups equally well. Take, for example, the personal data-protection measures embedded in the federal decennial census. Census data are "an all-important component of the estimation of population-level indicators for fertility, health, migration, and mortality" (Santos-Lazada, Howard, & Verdery 2020). Yet it has been argued that differential privacy data-protection methods (Dwork & Roth 2014) used in connection with the 2020 Census disadvantaged population group members affected by health disparities by significantly misestimating mortality rates for non-Hispanic Black people and Hispanics, compared to non-Hispanic white people (Santos-Lazada, Howard, & Verdery 2020).

Notwithstanding the inequitable distribution of privacy protections, and the fact that conditions of privacy can be exploited to conceal serious wrongdoing by institutions and individuals, I believe health data privacy merits the robust protection of law (Allen 2011). Some legal obligations of privacy, confidentiality, and data security already significantly bind physicians, nurses, social workers, mental health providers, pharmacists, hospitals, insurers, health data processors, health researchers, and public health officials (Allen 2021b). Health care privacy is a mandated priority of providers and biomedical researchers in the regulatory ecosystem that includes state confidentiality laws, the Health Insurance Portability and Accountability Act

of 1996 (HIPAA), and related regulations (Department of Health and Human Services 2000; 2003); the Genetic Information Nondiscrimination Act (United States 2009a); and the revised Common Rule (Department of Health and Human Services 2018b). Yet regulatory challenges to privacy protection abound, including permissive HIPAA rules and largely unregulated "nontraditional health-relevant data . . . in widespread commercial use" (Allen 2021a; McGraw & Mandl 2021). Individuals are vulnerable to unwanted, unintended, or harmful disclosures of health-related information on multiple fronts: social media, wearable health monitoring devices, health apps, direct-to-consumer health screening and DNA testing, genomic research, biobanking, big data algorithmic analytics, public health measures and surveillance, data breaches, law enforcement, national security, and the judicial process. If all Americans are to benefit, addressing these challenges must proceed with the aid of a new generation of well-designed, inclusive, and equitable laws not currently on the horizon.

Evolving Health Privacy Goals

What is most at stake in current debates over optimal regulatory approaches to health data privacy? Some might respond "barriers to efficient innovation," while others might say "human dignity." Because health data privacy is a particular kind of health privacy, answering this question—which is a question about values, goals, and practical challenges—usefully begins by taking a step back to gain perspective on how our contemporary concerns about health data privacy relate to traditional concerns about health privacy. Has something changed to make privacy less important?

Concerns about data as such are relatively new; concerns about health privacy are not. The legal regulation of health privacy, I argue, has always included concerns about who should have access to and control over information about individuals. Examples spanning the centuries suggest what is different, what has remained the same, and what we must do to help preserve health privacy into our collective futures.

I illustrate an evolution in how Americans have thought about health privacy using litigation examples spanning the 19th, 20th, and 21st centuries. The examples illustrate how our developing legal system has functioned to protect health-related privacy over time. Three distinguishable policy goals lay behind these examples: enforcing customary morality, constraining public policy in the interest of individual rights, and governing a complex health infrastructure.

Enforcing Customary Morality

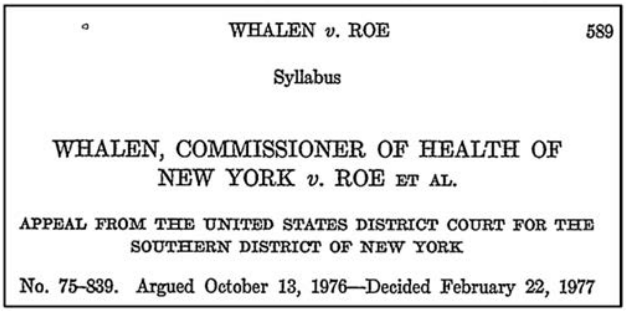
In the late 19th century, when preserving privacy in the context of medical care was a large concern, there was little formal legal regulation of health privacy. Yet courts in that period adjudicating personal injury cases recognized the importance of limiting access to medical encounters and the health-related information they generate in the interest of public enforcement of private morality. A Michigan state court case, *De May v. Roberts* (1881), illustrates the point. Mr. and Mrs. Roberts sued Dr. De May, a physician who came to their home to deliver their child, and Mr. Scattergood, a friend of the physician, who had accompanied him to the Roberts's home. Since by custom only physicians, midwives, and family members were morally acceptable witnesses to childbirth, Mr. and Mrs. Roberts had presumed Scattergood to be a medical colleague of Dr. De May. They later learned he was an "unprofessional young unmarried man" (*De May* 1881), only brought along to help carry the tired and sick doctor's belongings. Scattergood had entered and remained inside the Roberts' small house throughout a protracted labor and had even helped hold Mrs. Roberts steady during a paroxysm of pain. Ruling against De May and Scattergood on appeal, the Michigan high court asserted, "To the plaintiff the occasion was a most sacred one [and she] had a legal right to the privacy of her apartment at such a time" (*De May* 1881). The court affirmed that De May and Scattergood were liable for fraud and deceit. They had wrongfully exposed Mr. and Mrs. Roberts to the "shame and mortification" of having allowed Scattergood to hear, observe, and participate in intimacies of childbirth without knowledge of "his true character" (*De May* 1881).

A U.S. Supreme Court case, *Union Pacific v. Botsford* (1891), from the same period similarly elevates privacy and customary morality governing women's modesty above competing practical concerns. The question put to the Court in an appeal was whether a person alleging to have been injured on a train could be compelled "without his or her consent, to submit to a surgical examination as to the extent of the injury sued for” (*Union Pacific* 1891). Clara Botsford alleged that while a passenger in a negligently constructed sleeping car, the upper berth fell on her head, “rupturing the membranes of the brain and spinal cord and causing a concussion [and] permanent and increasing injuries” (*Union Pacific* 1891). Given the public interest in deterring fraud and protecting the assets of a business on which the public depends for transportation, one might suppose that persons seeking to recover money damages for alleged negligence could be compelled to provide medical evidence of an injury based on a physical examination. Yet the Court opined that “To compel anyone, and especially a woman, to lay bare the body, or to submit it to the touch of a stranger … is an indignity, an assault, and a trespass” (*Union Pacific* 1891). The Court observed in a shocked tone, “The inviolability of the person is as much invaded by a compulsory stripping and exposure as by a blow” (*Union Pacific* 1891).

These two 19th-century cases dealt with health information that can be wrongfully obtained by improper physical contact through the five senses. Contact with a patient in violation of customary morality potentially subjects a woman (and, if she is married, her spouse) to harms that include moral censure, diminished reputation, as well as shame and embarrassment. The injuries are material and dignitarian. When coercion (as in *Union Pacific*) or deception (as in *De May*) enters the picture, interference with autonomy and decisional privacy are implicated harms as well.

Constraining Public Policy in the Interests of Individual Rights

A few generations later, in the latter half of the twentieth century, the same concerns about autonomy and disclosure arose anew in connection with legally constraining the reach of public policies mandating governmental collection of health information. The U.S. Supreme Court’s decision in *Whalen v. Roe* (1977) is celebrated as its first major recognition of a constitutional privacy or liberty right in health information. In this case, anonymous patients, physicians, and professional groups sued the state of New York complaining that the equal protection and due process rights guaranteed in the Fourteenth Amendment were violated by a law requiring that the names and addresses of persons prescribed certain drugs be reported to the state and stored in a computer system. Specifically, the law required that physicians submit forms to



WHALEN *v.* ROE 589

Syllabus

WHALEN, COMMISSIONER OF HEALTH OF NEW YORK *v.* ROE ET AL.

APPEAL FROM THE UNITED STATES DISTRICT COURT FOR THE SOUTHERN DISTRICT OF NEW YORK

No. 75-839. Argued October 13, 1976—Decided February 22, 1977 the state reporting their prescription of specified "Schedule II" drugs—otherwise legal drugs with a high potential for abuse and illegal sale or distribution. Completed forms were to identify "the prescribing physician; the dispensing pharmacy; the drug and dosage; and the name, address, and age of the patient" (*Whalen* 1977). The law mandated that copies of the form be sent to the pharmacist and to the New York State Department of Health (NY-DOH) in Albany. Once at the NY-DOH, the forms were "sorted, coded, and logged" and the data "recorded on magnetic tapes for processing by a computer" (*Whalen* 1977).

Plaintiffs in the *Whalen* case claimed that the reporting of names was unnecessary and endangered health. Patients who feared "the misuse of the computerized data" and being stigmatized as "drug addicts" were refusing prescription medications (*Whalen* 1977). Finding the law sweepingly broad, a lower court had enjoined enforcement of provisions that required the reporting of individual names and addresses, holding that "the doctor-patient relationship intrudes on one of the zones of privacy accorded constitutional protection." New York appealed its loss to the Supreme Court and won.

The Supreme Court expressed awareness of "the threat to privacy implicit in the accumulation of vast amounts of personal information in computerized data banks or other massive government files" and affirmed that disclosures of private matters and interferences with autonomy are constitutional interests protected by the 14th Amendment (*Whalen* 1977). The Court decided the case by, in effect, balancing the state's rational interest in requiring disclosure of drug-prescribing data that identified the patient with a competing 14th Amendment interest in privacy. Finding in favor of the state, the Court stressed that New York had exercised its constitutional police powers through reasonable regulation to address a legitimate governmental interest in drug control; that "disclosures of private medical information to doctors, to hospital personnel, to insurance companies, and to public health agencies are often an essential part of modern medical practice"; and that the state had implemented sufficient administrative protections to make it unlikely that identifying information would ever be improperly disclosed. To this day, under the New York Codes, Rules, and Regulations, physicians and pharmacists remain subject to official prescription reporting requirements, and the state maintains a registry of individuals' prescription histories. Electronic storage of medical information, with security measures in place, has been normalized.

Governing a Complex Infrastructure of Health Institutions and Practices

The digitalization of health data by government, business, and technology companies has become the norm on a scale scarcely imaginable in the era of *Whalen*. How these data are collected, shared, and secured by hospitals, medical professionals, and insurance companies is closely regulated by HIPAA, with some private sector health data businesses falling outside HIPAA's scope (United States 1996; Allen 2021b). HIPAA ushered in a sweeping new legal landscape for health information privacy—replacing both the customary morality approach, presupposed by *Union Pacific*, and the privacy rights approach of *Whalen v. Roe*. Importantly, individuals do not have a private right of action under HIPAA; they may not bring a lawsuit on their own behalf when protections under the statute are violated.

HIPAA has been the major federal guardian of health privacy rights for 25 years and was America's first national health privacy statute. HIPAA authorized the U.S. Department of Health and Human Services (HHS) to develop what have come to be known as a privacy rule and a security rule (Department of Health and Human Services 2000; 2003). These regulations aim at protecting the confidentiality and electronic security of personal health information (PHI) and the electronic health or medical record (EHR or EMR).

HIPAA rules protect informational privacy in clinical and research settings. Stated rationales for health information privacy in the HIPAA context include freedom from discrimination by employers and health or life insurers, and protection from identity theft and fraud, along with traditional concerns about moral censure, reputation, stigma, shame, and embarrassment (Deapen 2006). Covered entities—health plans, health care clearinghouses, and most health care providers—must protect patients' identifiable health information from misuse and must limit sharing. With exceptions for emergencies and disasters, covered entities must obtain written authorization before using or disclosing identifiable health information for treatment, payment, health care operations, or commercial purposes. Transactions with business associates—persons or companies that partner with covered entities to perform health care functions—require HIPAA contracts that explicitly permit use or disclosure of health information.

HIPAA is a highly technical, complex response to an increasingly complex and digital data–dependent health economy. HIPAA and its regulations have been revised periodically in response to emerging issues, including genomics, medical informatics, cryptography, big data algorithmic analytics, wearable health devices, and telemedicine (Allen 2021b). The Genetic Information Nondiscrimination Act (GINA), which amended HIPAA to restrict the use of individual genetic data by health insurers and employers, reflects traditional privacy values. Other amendments stray far afield from traditional privacy values and concerns. Some make it easier to share health data for business and public health purposes without specific notice or the consent of the data subject. The Health Information Technology for Economic and Clinical Health (HITECH) Act (United States 2009b) promoted the use of EHRs and other information technology. An Omnibus Rule modified HIPAA, GINA, and the HITECH Act to improve their workability, effectiveness, and flexibility. This rule aimed to better balance individual rights with public health and medical research by, for example, continuing to allow access without specific patient authorization to limited datasets and deidentified patient information (Department of Health and Human Services 2013). In December 2020, HHS again proposed amending the privacy rule. The aims of the new rule included "improving information sharing for care coordination and case management for individuals; facilitating greater family and caregiver involvement in the care of individuals experiencing emergencies or health crises; enhancing flexibilities for disclosures in emergency or threatening circumstances, such as the opioid and COVID-19 public health emergencies; and reducing administrative burdens on HIPAA covered health care providers and health plans" (Allen 2021a).

HIPAA and its rules and amendments set the terms under which efficient access and disclosure take place. This is the virtual opposite of *Union Pacific or De May*—and a major shift from *Whalen*, which acknowledged an individual's right to privacy as a strong constraint on state access to health data. As vaunted rights to health information privacy grow more technical and qualified, it is hard to interpret HIPAA as functioning to protect a singular right to health data privacy at all.

The HHS Office for Civil Rights (OCR) is responsible for enforcing the privacy rule and the security rule. OCR has investigated national pharmacy chains, major medical centers, group health plans, hospital chains, and small provider offices. Most HIPAA violations involve unauthorized disclosures, inadequate record disposal, poor training, dishonesty, hacking, identity theft, or large data breaches (Yaraghi & Gopal 2018). Massive data breaches in the health care industry are common phenomena, potentially exposing sensitive personal and personal health information and facilitating identity theft and fraud. Following the logic of *Whalen*, one might conclude that alarming violations of data privacy rights are occurring with disturbing frequency.

Over the past decade, Anthem, the health care insurance industry giant, has been beset by multiple HIPAA enforcement actions, as well as private suits alleging negligence and breach of contract. As a result of a massive breach announced in 2015 and affecting the data of nearly 80 million individuals, Anthem agreed in 2017 to settle private class action lawsuits for $115 million, the largest-ever medical data breach settlement (*In re Anthem, Inc. Data Breach* 2018). Records containing personally identifiable information were breached, including medical identification numbers, names, addresses, employment information, and Social Security numbers. Anthem separately reached a $16 million settlement with HHS in an enforcement action alleging violations of HIPAA in the 2015 breach (Department of Health and Human Services 2018a). The Supreme Court upheld the laws at issue in *Whalen* because New York had implemented effective data security measures. Yet today, sensitive health data continue to be collected, even though data security measures frequently fail due to human error, software deficits, or malicious hacking.

What Do We Most Value?

Two narratives of health data privacy have competed in the marketplace of ideas. The older, fading narrative asserts that the privacy of health-related information is vital to an array of utilitarian and dignitarian ends, giving rise to the need for stringent ethical and legal measures to protect it. Privacy is paramount. The newer, ascendant narrative contends that innovative medical care, public health, research, and pharmaceutical development require vast quantities of data about individuals in order to effectively and efficiently diagnose, treat, prevent, and cure illness and disease. Securing access to patient data, hopefully in ways that are transparent, accountable, and protective of privacy, is paramount. Because the new narrative continues to hold up privacy as a value, the slippage in policy preference isn't always immediately obvious. But close examination reveals the change. Health privacy is not the preeminent concern it once was (although complying with HIPAA and other "privacy" regulations takes up a lot of time and energy). Instead, reasonable management of health enterprises dependent on data sharing is the preeminent concern. Privacy is at the table, but not the guest of honor.

The Old Narrative

The older narrative classifies some types of health information as especially sensitive and in need of protection. Versions of the fading narrative treated some information about an individual's health history as if it were especially sensitive, including information about infertility, sexually transmitted infections, abortions, heritable genetic conditions, and even cancer. Protecting genetic information and HIV/AIDS status information were at various times put forward as being of essential importance in the health privacy realm. The old narrative understands privacy, confidentiality, and data protection as the path to better health, asserting that individuals would avoid or delay medical care or fail to provide honest and complete information to medical professionals if they could not count on privacy and confidentiality. And while health privacy under the fading narrative was typically associated with the confidentiality of identifiable patient information, physical privacy in clinical and surgical settings (gowns, curtains, private examination, and hospital rooms) was a key societal expectation of the ethical delivery of health care and conduct of research.

In the 1990s, prevailing bioethical perspectives emphasized the importance of privacy and confidentiality in clinical, research, and health administrative settings (Dhir & Aggarwal 1998). Awareness of HIV/AIDS stigma was at a peak, and patient advocates and activists insisted that a basic and primary duty of clinicians and researchers was to protect patient privacy in treatment, research, and billing. Some states enacted special laws governing HIV/AIDS information. The Human Genome Project was in full swing in the early 1990s, when concern about genomic stigma and insurability flourished in policy circles (Allen 1997). Ethical, legal, and social perspectives stressed that genetic data required privacy, perhaps exceptional levels of privacy compared to other health-related data (Gostin & Hodge 1999). When Congress enacted HIPAA in 1996, the old narrative was still dominant, with the law calling for privacy and security rules. Over time, HIPAA rules have been amended to ease sharing, and it would be difficult to defend them as embodying the old narrative any longer.

Among the pressures on the older ideals of health privacy is the concern that socially useful, even lifesaving technological innovations could be thwarted by old-fashioned moralistic and individualistic privacy impediments held over from the 19th and 20th centuries. This critique is not new. Philosophers have argued in the face of feminist and progressive critiques that although the concept of privacy can be deployed for antisocial and inequitable purposes, privacy is not an inherently selfish or elite value (Allen 2011). A main function of privacy protections is to empower individuals to be more fit for their varied quotidian roles and social responsibilities in our inevitably communal world. Indeed, "Opportunities for individual forms of personal privacy make persons more fit for social participation and contribution to the pool of resources and assets available to all" (Allen 1988, p. 51). Dignity, self-determination, and well-being require opportunities for privacy and private choices, making privacy an aspect of the common good. Because it facilitates wellness and independent thought and action, privacy is vital for democratic life.

The New Narrative

In the third decade of the 21st century, a shift from emphasizing to deemphasizing privacy has become discernable. Under a new narrative, privacy interests are lightly addressed through data security; practices of informed consent, notice, and opt-out rights; and trust and transparency measures. The old and the new narratives both aim at improving health. The new narrative, which may favor large pharmaceutical and technology companies, research universities, and governments, offers sharing as the path to better health.

Examples of the new narrative are everywhere. The federal government's marketing of the All of Us Project, part of the Precision Medicine Initiative (Denny et al. 2019), promotes and relies on the new narrative, aggressively encouraging information sharing. The goal of enrolling a million diverse Americans in a voluntary program that will make their personal health and genomic data available to approved future researchers is premised on the idea that sharing health information is ethically responsible conduct and a public service. All of Us, like the Supreme Court in *Whalen v. Roe*, stresses that government's data privacy and security measures make unwanted data disclosures unlikely and data sharing, therefore, safe. Yet data breach and deidentification risks must give pause (Mandl & Perakslis 2021).

A new narrative is suggested in the mission of the Stanford University Center for Artificial Intelligence in Medicine and Imaging, whose co-director Matthew Lungren has said, “We want to double down on the idea that medical data is a public good, and that it should be open to the talents of researchers anywhere in the world” (Health Care IT News 2021). With applications of artificial intelligence and machine learning at a historic high point, the privacy of health data, and indeed any data, has come to be seen as a barrier to innovation. The older ideal promoting health data privacy is clashing with a newer ideal promoting health data sharing.

The new narrative was apparent in a 2019 National Academy of Medicine white paper, “Empowering 8 Billion Minds: Enabling Better Mental Health for All via the Ethical Adoption of Technologies” (Doraiswamy et al. 2019). The paper highlighted the value and growing availability of technology-based mental health interventions and urged decision-makers and stakeholders to “focus on transparency and security to build trust by shifting the discussion from one of data privacy to one of transparency (how and why data is used and by whom), control (by the individuals themselves) and security (to guard against any misuse)” (Doraiswamy et al. 2019).

The Public Is Wary and Off Guard

The success of the new narrative depends upon persuading the public that privacy is not important, or that they have more to gain than lose by voluntarily sharing data. Some people may jump at the idea of handing over health data, while others balk. Giving notice of data practices and obtaining informed consent to them has been widely adopted by the business sector as a fair and lawful way to honor privacy. Researchers are expected to rely on deidentification, anonymization, and pseudonymization to shield individual data privacy (McGraw & Mandl 2021). Yet reidentification is a risk, and neither HIPAA rules nor the recent

crop of state privacy laws in California, Colorado, Connecticut, Utah, and Virginia clearly and adequately regulate all personally identifiable information relevant to health (Yoo et al. 2018; McGraw & Mandl 2021). African Americans have not forgotten the Tuskegee syphilis experiments (Jones 1981), the Henrietta Lacks "immortal" cell-line story (Skloot 2010), or Philadelphia's Holmesburg Prison dermatology experiments on incarcerated Black men (Wilkie 2000). The same Public Health Service responsible for the Tuskegee experiment committed even worse offenses in Guatemala, where, in cooperation with local officials, they deliberately exposed vulnerable populations to sexually transmitted infections (Presidential Commission for the Study of Bioethical Issues 2011). Marginalized groups look to history and see red flags of disrespect and outright harm. Many people just want to be left alone.

The traditional emphasis in academic and policy discussions of privacy, informed consent, and choice grounded in individual freedom can neglect the ways in which health disparities and structural injustices contribute to poor health, constrain choice, and reduce opportunity (Obasogie & Darnovsky 2018). These background conditions, along with pro forma point-of-service privacy notices and signatures, mean that informed consent may be illusory. Moreover, the rise of big data and artificial intelligence in the digital economy has made it increasingly difficult for any individual, especially those from disadvantaged groups, to exercise meaningful control over the collection, manipulation, and use of medically related and other personal information about them (Waldman 2021; Velíz 2020; Allen 2016).

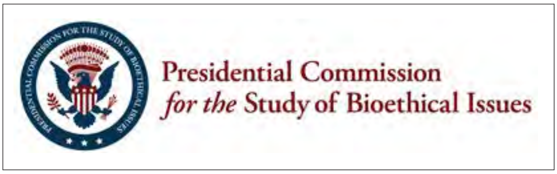
The reality today is that it is not merely drug companies, academic health researchers, and clinicians who seek and use health data. Consumers who eschew research participation or formal clinical care nonetheless feed the health data machine. They do so when they order health supplements from Amazon. They provide bits of health data "voluntarily" every time they search "flu" or "headache" or "rash" on Google, or log into a site such as breastcancer.com. Wellness and recreational platforms are extremely popular with consumers. Apple Watch apps, Fitbit, Peleton, and Noom apps are feeding health data to the platform economy. In addition, consumers use direct-to-consumer medical tests and the genetic tests offered by 23andMe or Ancestry.com, avoiding the realm of formal EHRs, but sharing health data nonetheless. On any given day, one individual may tell their selection of black-box platform apps how much they weigh, what they ate, how many steps they walked and where, and their heart rate, blood pressure levels, blood sugar, mood, and sleep and awake times.

With this panoptic data flowing out, how well does the typical consumer understand what happens to their data? Could data they would not directly give to a university researcher or the government end up in those hands anyway through sharing they unwittingly authorized? Consumers provide a great deal of information about matters as consequential as their entire genome and as minor as their blood pressure or heart rate at a certain time of day. Few consumers grasp the significance of all that sharing. Data analytics can reveal things about us, such as if we are pregnant or have postpartum depression, with tiny bits of seemingly unrelated information that we may wish to conceal or perhaps not even know ourselves (Chancellor et al. 2019).

How to Strike a Balance

Introducing a group of papers presented in London in 2002 at a conference, “Privacy and the Secondary Use of Data in Health Research,” organized by the Royal Society of Medicine in association with the Nuffield Trust, Nick Black posed the dilemma of privacy versus disclosure in health research: “Must privacy and medical confidentiality be compromised in data base research? Or must the pursuit of new knowledge be stifled because of privacy concerns? Can privacy and research be served at the same time?” (Black 2003, p. 1). If privacy and new knowledge are framed as being on equal footing, the only satisfying response to the perceived dilemma is to find a way to have a reasonable modicum of both privacy and disclosure. Having it both ways is often described through the metaphor of achieving a balance. Yet it is easier to agree that we need policies that balance competing priorities than to specify a set of practices that fit the bill. One person’s balance is another person’s skew. Seeking policy balance has unavoidable subjective and political dimensions. A balance from the vantage point of the old privacy narrative may entail giving privacy greater weight than a balance from the vantage point of the new narrative.

Calls for balance are ubiquitous in discussions of health data. For example, excited by the “new science of harnessing diverse streams of digital information to inform public health and policy,” data protection researchers argued for a “new balance between controls on collecting information and controls on how it is used” (Horvitz & Mulligan 2015, p. 255). The potential of this new science, which is not without its ethical risks (Chancellor et al. 2019), was suggested by research showing that accurately diagnosing postpartum depression in new mothers may be more easily achieved by analyzing new mothers' online activities and language on social media than by eliciting self-reports of how they are feeling. This kind of "digital disease detection" and "infodemiology" enabled by "machine learning presents new challenges for protecting individual privacy and ensuring fair use of data" (Horvitz & Mulligan 2015).



Reporting on the need to balance privacy and progress in connection with genome sequencing and biobanks, President Barack Obama's Commission for the Study of Bioethical Issues* recommended that national and state policymakers enact a consistent set of privacy protections governing how genomic data can be collected, stored, and shared (Presidential Commission for the Study of Bioethical Issues 2012). The commission's recommendations left it to policymakers to craft politically viable laws and regulations that would give privacy its due, without impeding biomedical progress. The commission's perspective on what that meant emphasized notice, informed consent, and, less predictably at the time, deliberative participation by stakeholders.

Going further than most toward describing what balance looks like, Allan Ford and W. Nicholson Price addressed the tension they perceive between privacy and the accessibility of data with a principled resolution (Ford & Price 2016) that moved beyond the notice and informed consent paradigm popularized 20 years prior. Black-box, big-data-reliant medicine could transform health care for good, they argued, but at a price. Health researchers need access to massive amounts of health information to develop black-box algorithms, putting patients at risk of privacy losses. Independent researchers need access to this same information to verify black-box algorithms, ensuring that they are accurate and unbiased, but at the risk of further privacy losses. Balancing this double-layer tension between accountability and privacy "is a key challenge in the development of black-box medicine" (Ford & Price 2016, p. 42).

Ford and Price (2016) named their solution "privacy-preserving accountability," a concept built on three pillars: (1) "researchers developing black-box algorithms should comply with substantive limitations on the collection, use, and disclosure of patient health information;" (2) "independent gatekeepers should oversee information sharing between those developing and verifying black-box algorithms;" and (3) robust information-security requirements should be imposed to prevent unintentional data breaches of patient information. Pillar one was consistent with notice

* The author was a member of the commission.

and informed consent as substantive limitations of law and policy, but allowed for the implementation of more exacting substantive limitations. Pillar three called for data security, a traditional fair information practice and legal expectation, but with an urgency arising out of the data breaches that numerous health-related institutions experienced in the 2000s. Pillar two was the most "modern" guideline, calling for "independent gatekeepers" to look out for the public's interests in a way that a self-interested business or software developer cannot. Independent gatekeeper accountability was not and still is not deeply entrenched in law or policy.

INPUT
BLACK BOX
OUTPUT

Ford and Price (2016) argued that rules premised on their three broad principles can help ensure that patients obtain the benefits of black-box medicine without sacrificing their privacy. One can only hope that policymakers feel sufficiently guided by high-level pillars and principles as they are left to specify the precise rules. Rules requiring transparency, encryption, notice and consent, and deidentification are among the usual measures offered to ensure that privacy is protected without standing in the way of research progress. However, experts may push policymakers for more in the future, such as the use of expert bodies, trusted intermediaries, privacy by design, and privacy-enhancing technologies (Jordan, Fontaine, & Hendricks-Stirrup 2022).

Providing more specificity than Ford and Price, McGraw and Mandl proposed five guidelines to strike a "balance between protecting patients and making data available to improve health and health care" (McGraw & Mandl 2021). Their guidelines centered around "HIPAA's framework and FTC consumer privacy recommendations." This choice of framing could concern privacy advocates who see HIPAA as lacking a strong ethos of individual privacy protection, and who understand the U.S. Federal Trade Commission (FTC) as pushing a traditional fair-information practice "notice and consent" framework and as limited in its enforcement powers by its unfair trade jurisdiction.

McGraw and Mandl's five guidelines are (1) increased transparency and choice for consumers; (2) limitations on how health data can be collected, used, and disclosed instead of relying only on consent; (3) mechanisms to assure beneficial uses of health-relevant data (e.g., independent data ethics boards, health trusts, impact assessments, and accountable data custodians); (4) strengthened remedies for harms incurred from malevolent uses of health data; and (5) accountability for uses of deidentified data. While guidelines 1, 2, 4, and 5 appear to weigh in favor of individual privacy, guideline 3 weighs on the side of disclosure by placing decisions about permissible health data–sharing practice in the hands of experts and third-party intermediaries who are to "assure" that data are available for beneficial purposes. Does assuring that data are available for useful purposes mean that privacy will never be a reason to keep data out of the hands of someone who can articulate a beneficial use? Is it a sign that the new narrative has won out over the old when balancing privacy and disclosure includes the assumption that beneficial public purposes override assertions of privacy?

Minority Voices in the Balance

A different approach comes from bioethicists who recognize privacy and disclosure as the horns of a dilemma, but who reject the notion that balancing interests is how we can resolve that dilemma. Ronald Bayer and Amy Fairchild addressed the "tension between the public's need for knowledge as a foundation for rational policymaking and the centrality of norms of privacy for democratic societies" in a discussion of name-based (nonanonymous) public health surveillance systems for HIV/AIDS, cancer, and drug abuse (Bayer & Fairchild 2010). Based on the history they recount, patient-group activists sometimes advocated against privacy and in favor of disclosure mandates as the old narrative was fading. For example, in 2010, "The attention of AIDS activists had turned from the privacy-informed agenda of the epidemic's early years to the challenge of assuring that those with HIV, an increasing proportion of whom were black and Hispanic, had access to the expensive long-term care they needed for survival." (Bayer & Fairchild 2010, p. 914). Bayer and Fairchild ultimately rejected the dream of policies that purport to balance privacy and disclosure in public health surveillance without requiring tradeoffs. They saw "the tension between privacy and the need to know on the part of public health agencies as enduring even if it is not always expressed in bitter controversy" (Bayer & Fairchild 2010, p. 906). An open recognition of ongoing tensions "hold[s] out the prospect for recognizing both the claims of privacy and public health" (Bayer & Fairchild 2010, p. 914). Democracy may indeed, as they suggest, require "an ongoing effort to negotiate and renegotiate the boundaries between privacy, societies' 'limiting principle' and public health … which is a guardian against disease and suffering" (Bayer & Fairchild 2010, p. 926).

Bayer and Fairchild raise an important question about whether health data privacy is the friend or foe of African Americans and other people of color in marginalized population groups affected by health disparities and injustice. Privacy resources can translate into valuable social, political, and economic power (Velíz 2020). While privacy is extremely important, acquiring essential care and medication may be more important to a community than better informational privacy. There is little point in having excellent privacy rights and a confidential and secured medical diagnosis if lack of access to medical treatment is a death sentence. Similarly, an individual may find it more important to speak out about their health condition, or to call attention to the need for more inclusive research, than to keep that condition secret. The preferences of affected populations in terms of favoring more privacy or more disclosure merit respect as part of the quest for health equity. A risk with the new narrative of health privacy is that in its most extreme forms it assumes that more information sharing is the right starting point when seeking balance.

Conclusion

Society has not outgrown the need for strong privacy protections. An older conception of health data privacy reflected in 19th- and 20th-century examples assumed the existence of fundamental moral and legal rights grounded in values that included dignity, intimacy, and well-being.

A new, technocratic vision of health data privacy has taken hold in the 21st century, whereby privacy interests are considered among many others, without pride of place. This new vision and associated narrative is discernible in HIPAA's evolving regulations, which rest only lightly on the old values. Digital technology innovations and big data black-box analyses have helped fuel a shift from a strong policy preference for privacy and confidentiality to a strong preference for sharing and disclosure. A preference for disclosure is even visible in some efforts to balance privacy and disclosure.

The challenge ahead is to figure out how to give both health innovation and privacy their due. The framing of balancing data sharing and data privacy keeps privacy in the picture, but care must be taken to avoid the kind of balance that guarantees that an entity seeking data for a legitimate purpose in the public interest can rarely, if ever, be told "no." Policymakers are helped by the principles and guardrails offered by experts to guide their work. Those with the power to decide and legislate must assume their responsibility to reengage and perhaps reassert privacy values as they seek to design and regulate our futures. And as they craft a new balance, let us hope they keep their ears to the ground, and not just their fingers in the air to feel the direction of the wind.

Anita L. Allen, J.D., Ph.D., is the Henry R. Silverman Professor of Law and a professor of philosophy at the University of Pennsylvania and an expert on the philosophical dimensions of privacy and data protection law, ethics, bioethics, legal philosophy, women's rights, and diversity in higher education. She served as Penn's vice provost for faculty (2013–2020), on President Obama's Presidential Commission for the Study of Bioethical Issues (2010–2017), and on numerous boards, including the National Constitution Center, the Future of Privacy Forum and the Electronic Privacy Information Center, whose Lifetime Achievement Award she has received. Allen is author of more than 120 articles and chapters as well as books that include *Unpopular Privacy: What Must We Hide* (2011), *Privacy Law and Society* (2017), and *The New Ethics: A Guided Tour of the 21st-Century Moral Landscape* (2004).

References

Allen, A. L. (1997). Genetic privacy: Emerging concepts and values. In M. A. Rothstein (Ed.), *Genetic secrets: Protecting privacy and confidentiality in the genetic era* (pp. 31–59). Yale University Press.

Allen, A. L. (1988). *Uneasy access: Privacy for women in a free society.* Rowman and Littlefield.

Allen, A. L. (2000). Gender and privacy in cyberspace. *Stanford Law Review*, 52(5), 1175–1200.

Allen, A. L. (2003). *Privacy isn't everything: Feminist reflections on personal accountability.* Rowman and Littlefield.

Allen, A. L. (2007). Face to face with “it”: And other neglected contexts of health privacy. *American Philosophical Society*, 151(3), 300–308.

Allen, A. L. (2011). *Unpopular privacy: What must we hide*. Oxford University Press.

Allen, A. L. (2014). Privacy in health care. In B. Jennings (Ed.), *Encyclopedia of bioethics* (4th ed., pp. 2064–2073). Macmillan Reference Books.

Allen, A. L. (2015). Compliance limited health privacy laws. In B. Roessler and D. Mokrosinska (Eds.), *Social dimensions of privacy* (pp. 261–277). Cambridge University Press.

Allen, A. L. (2016). Protecting one’s own privacy in a big data economy. *Harvard Law Review Forum,* 130, 71–80.

Allen, A. L. (2021a, June 10). HIPAA at 25—A work in progress. *New England Journal of Medicine,* 384(23), 2169–2171. doi: 10.1056/NEJMp2100900. Epub June 5, 2021. PMID: 34110114

Allen, A. L. (2021b). Privacy and medicine. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Spring 2021 ed.). https://plato.stanford.edu/archives/spr2021/entries/privacy-medicine/

Allen, A. L. (2022, February 20). Dismantling the Black Opticon: Privacy, race equity, and online data-protection reform. *Yale Law Journal Forum*, 131, 908–958.

Allen, A. L., & Rotenberg, M. (2016). *Privacy law and society* (3rd ed.). West Academic Publishing.

Arnett, C. (2020). Race, surveillance, resistance. *Ohio State Law Journal,* 81, 1103–1142.

Bayer, R., & Fairchild, A. L. (2010). When worlds collide: Health surveillance, privacy, and public policy. *Social Research,* 77(3), 905–928. https://www.jstor.org/stable/40972299

Black, N. (2003). Privacy, data, and health research. *Journal of Health Services Research and Policy*, 8(Supplement 1): 1.

Bridges, K. M. (2017). *The poverty of privacy rights*. Stanford Law Books.

Chancellor, S., Birnbaum, M. L., Caine, E. D., Silenzio, V. M. B., & Choudhury, M. D. (2019). A taxonomy of ethical tensions in inferring mental health states from social media. *Proceedings of Conference on Fairness, Accountability and Transparency 19*. https://doi.org/10.1145/3287560.3287587

Citron, D. K. (2014). *Hate crimes in cyberspace*. Harvard University Press.

Deapen, D. (2006). Cancer surveillance and information: Balancing public health with privacy and confidentiality concerns (United States). *Cancer Causes and Control, 17*(5), 633–637. http://www.jstor.org/stable/29736504

De May v. Roberts, 46 Mich. 160, 9 N.W. 146 (1881).

Denny, J. C., Rutter, J. L., Goldstein, D. B., Philippakis, A., Smoller, J. W., Jenkins, G., & Dishman, E. (2019). The “All of Us” research program. *New England Journal of Medicine, 381*(7), 668–676. doi: 10.1056/NEJMsr1809937. PMID: 31412182; PMCID: PMC8291101

Department of Health and Human Services. (2000, December 28). Health insurance reform: Standards for privacy of individually identifiable health information; Final Rule (HIPAA Privacy Rule), 65 Fed. Reg. 82,462 (codified at 45 C.F.R. § 164.500 et seq.).

Department of Health and Human Services. (2003). Health insurance reform: Security standards; Final Rule (HIPAA Security Rule), 45 C.F.R. Parts 160, 162, and 164.

Department of Health and Human Services. (2013, January 25). Modifications to the HIPAA privacy, security, and enforcement rules under the Health Information Technology for Economic and Clinical Health Act, and the Genetic Information Nondiscrimination Act; other modifications to the HIPAA Rules. Final Rule (2013 Omnibus Rule), 78 Fed. Reg. 5,566.

Department of Health and Human Services. (2018a, October 15). Anthem pays OCR $16 million in record HIPAA settlement following largest U.S. health data breach in history.

Department of Health and Human Services. (2018b). Policy for human subject research protections subpart A ("The Common Rule"). Title 45 C.F.R. 46 (Public Welfare) Subparts A, B, C, and D.

Dhir, K. S., & Aggarwal, S. S. (1998). Confidentiality of electronic medical records in the managed care environment. *Journal of Health and Human Services Administration, 21*(1), 80–91. http://www.jstor.org/stable/41426756

Doraiswamy, P. M., London, E., Varnum, P., Harvey, B., Saxena, S., Tottman, S., Campbell, P., Ibáñez, A. F., Manji, H., Al Olama, M. A. A. S., Chou, I., Herrman, H., Jeong, S. J.-S., Le, T., Montojo, C., Revel, B., Rommelfanger, K. S., Stix, C., Thakor, N., Chow, K. H.-M., Welchman, A. E., & Candeias, V. (2019). Empowering 8 billion minds: Enabling better mental health for all via the ethical Adoption of Technologies. NAM Perspectives Discussion Paper, National Academy of Medicine, Washington, D.C. https://doi.org/10.31478/201910b

Dwork, C., & Roth, A. (2014). The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science,* 9, 211–407.

Ford, R. A., & Price, W. N., II. (2016). Privacy and accountability in black-box medicine. *Michigan Telecommunications and Technology Law Review,* 23, 1–43.

Friedman, A. B., Pathmanabhan, C., Glicksman, A., Demiris, G., Cappola, A. R., & McCoy, M. S. (2022, January–December). Addressing online health privacy risks for older adults: A perspective on ethical considerations and recommendations. *Gerontology and Geriatric Medicine*, 8. https://doi.org/10.1177/23337214221095705

Garrison, N. A. (2013). Genomic justice for Native Americans: Impact of the Havasupai case on genetic research. *Science, Technology and Human Values,* 38(2), 201–223. https://doi.org/10.1177/0162243912470009

Gostin, L. O., & Hodge, J. G. (1999). Genetic privacy and the law: An end to genetics exceptionalism. *Jurimetrics, 40*(1), 21–58. http://www.jstor.org/stable/29762629

Harris, J. E. (2021). Taking disability public. *University of Pennsylvania Law Review, 169*, 1681–1749.

Health Care IT News. (2021, August 3). Stanford aims for global data access for AI research. https://www.healthcareitnews.com/ai-powered-healthcare/stanford-aims-global-data-access-ai-research

Horvitz, E., & Mulligan, D. (2015, July 15). Data, privacy, and the greater good. *Science, 349*(6245), 253–255.

In re Anthem, Inc. Data Breach. Litig., No. 15-MD-02617-LHK, 2016 U.S. Dist. LEXIS 70594 (N.D. Cal. Aug. 17, 2018).

Jones, J. H. (1981). *Bad blood: The Tuskegee syphilis experiment.* Free Press.

Jordan, S., Fontaine, C., & Hendricks-Sturrup, R. (2022). Selecting privacy-enhancing technologies for managing health data use. *Frontiers in Public Health, 10,* 1–13. https://www.frontiersin.org/articles/10.3389/fpubh.2022.814163/full

Mandl, K. D., & Perakslis, E. D. (2021, June 10). HIPAA and the leak of "deidentified" EHR data. *New England Journal of Medicine, 384*(23), 2171–2173. doi: 10.1056/NEJMp2102616

McGraw, D., & Mandl, K. D. (2021). Privacy protections to encourage use of health-relevant digital data in a learning health system. *npj Digital Medicine, 4*(2). https://doi.org/10.1038/s41746-020-00362-8

Obasogie, O. K., & Darnovsky, M. (2018). *Beyond bioethics: Toward a new biopolitics.* University of California Press.

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science, 366*(6464), 447–453.

Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information.* Harvard University Press.

Presidential Commission for the Study of Bioethical Issues. (2011). *Ethically impossible: STD research in Guatemala from 1946 to 1948.* Department of Health and Human Services.

Presidential Commission for the Study of Bioethical Issues. (2012). *Privacy and progress in whole genome sequencing.* Department of Health and Human Services.

Romoser, J. (2018). Rollback of privacy law would ease sharing of drug-abuse records. *InsideHealthPolicy.Com's FDA Week, 24*(4), 13–14. https://www.jstor.org/stable/27043936

Santos-Lozada, A. R., Howard, J. T., & Verdery, A. M. (2020, May 28). How differential privacy will affect our understanding of health disparities in the United States. *Proceedings of the National Academy of Sciences, 117*(24), 13405–13412. https://doi.org/10.1073/pnas.2003714117

Skinner-Thompson, S. (2020). *Privacy at the margins.* Cambridge University Press.

Skloot, R. (2010). *The immoral life of Henrietta Lacks.* Penguin Random House.

Union Pacific v. Botsford, 141 U.S. 250 (1891).

United States. (1996). Health Insurance Portability and Accountability Act (HIPAA) of 1996. Pub. L. 104-191, Title II, §§ 261, 264(a)–(b), 110 Stat. 1936, 2021, 2033.

United States. (2009a). Genetic Information Nondiscrimination Act (GINA) of 2008. Pub. L. 110-233, 122 Stat. 881.

United States. (2009b). Health Information Technology for Economic and Clinical Health (HITECH) Act, Pub. L. No. 111-5, Title XIII, 123 Stat. 115, 226.

Véliz, C. (2020). *Privacy is power: How and why you should take back your privacy.* Penguin.

Waldman, A. E. (2021). *Industry unbound: The inside story of privacy, data, and corporate power.* Cambridge University Press.

Whalen v. Roe, 429 U.S. 589 (1977).

Wilkie, T. (2000). Acres of skin: Human experiments at Holmesburg Prison. A true story of abuse and exploitation in the name of medical science. *Medical History, 44*(1), 132–133.

Yaraghi, N., & Gopal, R. D. (2018). The role of HIPAA Omnibus Rules in reducing the frequency of medical data breaches: Insights from an empirical study. *The Milbank Quarterly, 96*(1), 144–166. https://doi.org/10.1111/1468-0009.12314

Yoo, J. S., Thaler, A., Sweeney, L., & Zang, J. (2018, October 8). Risks to patient privacy: A re-identification of patients in Maine and Vermont statewide hospital data. *Technology Science,* 2018100901. https://techscience.org/a/2018100901/

"It is time to shift from the current safe harbor provision to a robust deidentification approach. It is more appropriate to define the categories of data that can safely be included in deidentified data rather than to base the safe harbor on categories that must be excluded."

– BRADLEY MALIN, PH.D.

Deidentification to Enhance Health Data Sharing

Bradley Malin, Ph.D.

Introduction

Clinical care generates a large amount of data related to an individual's health that can serve as the basis for biomedical research. These data include images, such as MRIs and CT scans; laboratory results; molecular information, such as genomic sequences; and natural language descriptions about a patient's history, progress, and guidance. Billing systems contain additional demographic information that reflects or is relevant to an individual's health. The volume of data is increasing rapidly as technology advances, allowing greater image precision (Perkel 2019), and as prices for technologies such as whole genome sequencing plummet (National Human Genome Research Institute 2021). Federal policies have spurred the adoption of health information technologies, placing much of these data into electronic health record (EHR) systems (Colicchio, Cimino, & Del Fiol 2019).

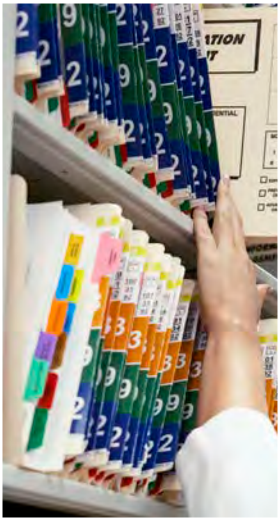
Data collected and documented in EHRs and related systems present growing opportunities for research (Wei & Denny 2015; Weiskopf et al. 2017; Williams et al. 2017). EHR-based data enable research to be conducted at large scale with real-world evidence gathered beyond clinical trials (Corrigan-Curay, Sacks, & Woodcock 2018). Numerous examples of such investigations exist. The findings published by the National Institutes of Health (NIH)-sponsored Electronic Medical Records and Genomics (eMERGE) network over the past 15 years demonstrate how EHR data can be standardized to support research investigations into numerous diseases, including rheumatoid arthritis, type-2 diabetes, breast cancer, and colorectal cancer (eMERGE 2021). More recently, the NIH-sponsored National COVID Cohort Collaborative (N3C) has shown how EHR data can support investigations into emerging infectious diseases, such as COVID-19 (Haendel et al. 2021).

Deidentification is a mechanism that was developed to make it easier to share health data. This paper provides a summary of what deidentification means and how it fits within existing data privacy laws. It then discusses the use of deidentification to support health data sharing and analytics and the potential for misuse, which can garner patient distrust and undermine the social benefits of data analysis. The paper concludes with a discussion of opportunities to strengthen regulatory frameworks for deidentification to maximize its benefits and minimize risks.

Deidentification as a Basis for Data Use

HIPAA Framework

Various federal laws and regulations, most notably the Health Insurance Portability and Accountability Act (HIPAA) of 1996, protect data generated in health care settings (United States 1996). HIPAA provides several mechanisms for using data.



Data can be used for their initial intended purpose, called primary use, with the patient's consent, which is typically provided through a HIPAA authorization form that indicates why the data will be disclosed. Data can be used in fully identified form for quality assessment and improvement within the institution where they were collected without informing the patient. As well, there are exceptional circumstances under which fully identified health data can be shared without the patient's consent, such as in response to court orders, court-ordered warrants, subpoenas, and administrative requests.

Data reuse, or secondary use, is the use of patient data for purposes beyond those for which explicit consent has been given. Data can be reused in a fully identified form for research purposes provided through a waiver of consent from an Institutional Review Board (IRB) if it is shown that it is impractical to obtain consent from the corresponding patients and the planned research poses minimal risk to the patient.

Data can also be reused without consent if they are stripped of certain identifying elements and the data recipient agrees to certain provisions. This is done by creating what is called a limited dataset, whereby 16 categories of information such as patient names, complete addresses, and medical record numbers are removed. The recipient of the data must enter into a contractually binding data-use agreement in which they indicate that they will not attempt to reidentify the corresponding patients and will not use the data for purposes beyond those indicated in the agreement. This is the mechanism by which N3C solicited EHR data from health care organizations across the United States.

This paper focuses on data reuse based upon deidentification. Deidentification is the transformation of personal data into a format that neither directly identifies, nor includes information that can be used to identify, the individual represented by the data. According to HIPAA, deidentified health data are no longer considered individually identifiable health information, meaning they are no longer covered by HIPAA's provisions, although they may be protected by other state and federal laws. HIPAA explicitly defines the steps necessary to convert personal data into deidentified data.

How Do Data Become Deidentified?

There are two mechanisms for deidentification under HIPAA: safe harbor and expert determination. Safe harbor extends the provisions of the limited dataset, eliminating additional categories of data, including geographic areas of residence with fewer than 20,000 individuals, ages over 89, and dates of events more precise than a one-year range. The organization sharing the data must further attest to the fact that it has no actual knowledge that the remaining information could be used to reidentify the corresponding patient.

While straightforward to implement, the safe harbor provisions can impede research where detailed geographic information, dates of events, or specific age ranges (e.g., an elderly population or neonates) are important.

Recognizing the limitations of the safe harbor option, HIPAA allows for an alternative that is typically referred to as expert determination. This provision is satisfied when:

> *A person with appropriate knowledge of and experience with generally accepted statistical and scientific principles and methods for rendering information not individually identifiable: (i) Applying such principles and methods, determines that the risk is very small that the information could be used, alone or in combination with other reasonably available information, by an anticipated recipient to identify an individual who is a subject of the information; and (ii) Documents the methods and results of the analysis that justify such determination (Office for Civil Rights 2013).*

The expert determination provision allows for flexibility that can facilitate important analyses. For example, the age of patients over age 89 can be retained as long as the patients' identities are sufficiently protected with respect to the other data that are shared (Malin, Benitez, & Masys 2011). Similarly, event dates may be retained or shifted by random offsets to capture the time between events (Hripcsak et al. 2016).

Expert determination also takes into consideration the risk related to how and with whom the data will be shared. For example, there is a different risk when publishing data on a publicly accessible web page than there is when transferring data to a researcher at an academic medical center. The transfer of a deidentified dataset can be combined with a contractual data-use agreement to further reduce risk, a strategy that various data-sharing endeavors have adopted. One example is the Medical Information Mart for Intensive Care (MIMIC)-III database from Beth Israel Deaconess Medical Center (Johnson, Pollard, & Mark 2016). Expert determination acknowledges that some level of risk is inherent in any data-sharing activity, but the risk permissible for data to receive a deidentification designation should be "very small."

The flexibility of expert determination brings with it certain ambiguities and challenges in translating legal requirements into practice. The Office for Civil Rights (OCR) in the U.S. Department of Health and Human Services (HHS), which is responsible for oversight of the HIPAA Privacy Rule, issued guidance in 2013 on how to ensure deidentification (Office for Civil Rights 2013). However, OCR stopped short of providing crucial clarifications, such as what is considered a "very small" risk or what credentials are needed to be a "person with appropriate knowledge and experience."

Benefits of Deidentification

Deidentification has several benefits for health data sharing and analytics.

Increased Data Sharing

Deidentification increases data sharing because it avoids the barriers present under the HIPAA-identified data and limited dataset provisions described above. As noted, sharing identified or limited datasets requires review from IRBs and creating and negotiating data-use agreements (DUAs). Because IRBs, DUAs, and organizational practices more generally are not standardized across institutions (Seykora et al. 2021), different organizations may request different contractual terms in their use agreements. The result is that recipients of the data must review each contract and determine whether they can meet its requirements, which add significant time and cost and make data sharing less likely.

In addition, deidentification can influence how breaches are reported. If data are not deidentified, an organization must notify the secretary of HHS if it discovers a breach of protected health information. If the breach affects 500 or more individuals, the breach must be reported without unreasonable delay, and no later than 60 days from discovery. If the breach affects fewer than 500 individuals, it must be reported within 60 days of the end of the calendar year in which it occurred. By contrast, deidentified data are not subject to breach reporting requirements under the HIPAA Security Rule. While this reduces the burden on organizations, it also lacks transparency and accountability, as discussed below.

Mitigation of Consent Bias

Data obtained through the patient consent process may not represent the patient population as a whole, introducing bias that can undermine the value of research involved. Deidentified data avoid this problem.

Health systems hold large amounts of HIPAA-protected data in their EHR and accounting systems that the patient has not consented to share for purposes beyond initial treatment, payment, or operations. It is extremely difficult to solicit consent

from individuals to share those data. Moreover, when consent is obtained, the disparity between consenters and nonconsenters is significant (El Emam et al. 2013). Research has demonstrated that these groups differ in demographic and socioeconomic characteristics, resulting in biased datasets (e.g., Wendler et al. 2005; Donkin et al. 2012).

These differences can exist for a number of reasons, including a failure to understand the request, distrust in the organization making the request (or distrust in the expected downstream organizations that may receive data), or simply a personal perspective that data should not be reused for other purposes. Regardless of the reasons, relying on individual consent results in datasets that are unlikely to be fully representative.

Data Linkages

The value of any dataset can be expanded if information about individual patients can be linked to data from other sources. To give a simple example, a patient's address can be linked to information about the community in which the patient lives, which can be useful for understanding the many factors that might affect the patient's health. Linking an individual to enrollment files for social programs can help determine the health effects of program participation.

One might imagine that deidentified data are less useful for research because data linkages are impossible, but this is not the case. While the goal of deidentification is to create data that have a low risk of reidentification, HIPAA permits deidentified data to include a "reidentification code." This is a code assigned by the covered entity that allows deidentified information to be reidentified by that entity. The code must not allow anyone else to identify the patient, and the data must not be disclosed to others.

Such a code is called a pseudonym.

The pseudonym makes it possible to integrate resources that create a more robust picture of a population and can substitute for the absence of a universal health ID in the United States. A pseudonym is often based on a combination of unique data that can include some combination of the person's Social Security number, name, ZIP code of residence, and gender.

An interesting side benefit of creating pseudonyms is that they can aid in eliminating duplicate records that bias study results. For example, a variation of this strategy was applied across a set of hospitals in Chicago. The analysis showed that failure to account for patients presenting at multiple health institutions could inflate disease prevalence estimates by as much as 28 percent for conditions such as asthma and diabetes (Kho et al. 2015).

In addition to reducing duplication and overcounting, linking records also provides a more complete picture of a patient's exposures and experiences. This is notable given the increasing recognition that many factors beyond the traditional biomedical domain influence an individual's health. The National Academies (Institute of Medicine 2015) and others have identified social determinants of health, such as educational attainment, income, and living situation, as important variables, yet we are a long way away from routinely collecting such information in the health care domain (Chen, Tan, & Padman 2020). Pseudonyms that link deidentified data are a tool for learning more.

Deficiencies of Deidentification

Although deidentification creates an opportunity for data sharing, it occurs in an unregulated environment, with little oversight, which can have several negative consequences.

Risk of Reidentification

Linking data through pseudonyms can raise the risk that a patient could be reidentified. That outcome, referred to as the mosaic effect, is the increased risk that transpires when multiple disparate pieces of data about an individual are linked (Office of Management and Budget 2013). Pseudonyms can provide a powerful solution to linking data from disparate resources, but care must be taken to confirm that the resulting integrated dataset does not raise the risk of reidentification above a very low threshold. By addressing this potential problem from the outset, the U.S. Centers for Disease Control and Prevention enabled the dissemination of deidentified versions of COVID-19 case surveillance data, which it continues to release to this day (Lee et al. 2021).

Ethics

Informed consent, the gold standard for collecting, using, and disclosing information, is a form of individual autonomy. Because deidentified data can be used without a patient's consent, a number of ethical issues surround its use for research (Caplan & Batra 2019).

Lack of Transparency

Under HIPAA, patients can request an accounting of their health data disclosures, allowing them to learn about the types of organizations that have access to their information and, at times, how it is used. This does not necessarily mean that a patient can prevent the transfer of information about them, but it does allow them to learn more about the practices of a covered entity and its business associates. In theory, a patient can choose to seek health care services from providers with data-handling practices in line with their expectations.

Once data are deidentified, however, patients are no longer entitled to such an accounting. Deidentification effectively severs the relationship between individual patients and their data. This prevents patients from learning that data about them have been deidentified or shared, even if a breach occurs.

Lack of Accountability

Since deidentified data are no longer subject to HIPAA, there is no clear liability for a breach. Even deidentified data pose a risk of reidentification, albeit a very small one, yet there is no clear legal path under HIPAA to hold the responsible party accountable.

This need not be the case. Examples of extending legal protections can be found in other environments where the notion of deidentification has been adopted. For instance, the five states that have adopted laws over the past several years to protect consumer data (California, Connecticut, Colorado, Utah, and Virginia) all include a deidentification provision. However, unlike HIPAA, these laws retain restrictions on the deidentified data. For instance, in the California Consumer Protection Act, deidentification is realized when four criteria are satisfied: (1) the information cannot reasonably identify, relate to, describe, be capable of being associated with, or be linked, directly or indirectly, to a particular customer; (2) the recipient must implement technical safeguards and business processes that prohibit reidentification; (3) the recipient must have implemented business processes to prevent inadvertent release of the deidentified data; and (4) the recipient must not make any attempt to reidentify the information. Provisions like these retain accountability for appropriate use of deidentified data.

Commoditization

Another concern with deidentified health data is that, since they are no longer regulated by HIPAA, they can be commoditized without oversight. An organization holding the data can profit from the sale of deidentified data and create markets for their sale (or auction). Indeed, there is already a billion-dollar market for deidentified health data. This was recently apparent when IBM announced that it would be selling its Watson Health group, which includes the Marketscan® research datasets, which IBM bills as “one of the longest-running and largest collections of proprietary deidentified claims data for privately and publicly insured people in the US” (IBM 2022).

The notion that a person’s data can be commoditized without their knowledge or consent may strike some as inconsistent with principles of privacy. The absence of regulation over this activity may undermine public confidence in the health data enterprise.

Viability

Deidentification at scale is viable, as the Vanderbilt University Medical Center (VUMC) has demonstrated. In 2007, VUMC began a process to deidentify all data (approximately 2.5 million medical records) into a repository called the Synthetic Derivative, making it available for research (Danciu et al. 2014). This data repository includes both structured data, such as diagnosis and procedure codes, and natural language text that has been subject to artificial intelligence methods to detect potential identifiers. Similar resources have recently been developed by other academic medical centers, including the Mayo Clinic (Anastasijevic 2020).

The Vanderbilt data repository was subsequently combined with deidentified genomic data extracted from blood left over from routine laboratory testing. This combined data repository, known as BioVU (Roden et al. 2008) has been used by more than 1,000 researchers to support numerous projects that were sponsored internally or by the NIH and other funding agencies, including the eMERGE network.

One caveat should be recognized here. Unlike the Synthetic Derivative, BioVU is a consented environment. This reflects the NIH genomic data-sharing policy, which requires NIH-sponsored research with genome sequences to obtain consent from the individuals to whom the data corresponds (National Institutes of Health 2014). Such consent requirements may make it impossible to use deidentification as a tool to reduce bias in datasets.

The pseudonym process and privacy preserving record linkage is also happening at scale. A growing number of companies are selling software systems to support these services, including Datavant, Symphony Health (through its Synoma software), and HealthVerity (through its Census software).* Moreover, these companies are creating ecosystems, such as Datavant Switchboard (May 2021) and HealthVerity MarketPlace (HealthVerity 2022), which enable disparate resources to be linked through, for example, medical claims from one data provider and EHR data from another.

* The author discloses having received benefits from Datavant and HealthVerity. The former purchased a company, Health Data Link, for which the author served on the board. The latter paid the author for assistance in developing its Census software tool.

Open Issues and Challenges

Weak Financial Incentives

Since deidentified data are no longer considered personal health information, they can be reused and disseminated with minimal restrictions. While that removes one barrier to data sharing, certain financial incentives remain to discourage it.

For example, consider a health care organization pursuing a $1 million NIH grant for a study involving deidentified health data. Such funding is often tied to the stipulation that the data be made accessible to others who wish to study them. That same health care organization could likely sell a private license to use the data to a life sciences or pharmaceutical company, perhaps at a price of $300,000. If the health care organization believes they can sell four or more such licenses, it is in their financial interest to license the data (thereby keeping it private) rather than obtain the NIH grant, despite its social benefit of making the data available publicly. This raises the question of how to structure or regulate the market for data in a manner that is consistent with maximizing the health benefits of analyzing deidentified data.

Improved Guidance

As mentioned earlier, the Office for Civil Rights has issued guidance on how to deidentify data in accordance with HIPAA, but it failed to specifically define key elements of the process. This is problematic because different deidentification experts will have different interpretations of terms like “very small” and “anticipated recipient.”

Agreement on what these terms mean and how they should be applied would remove this ambiguity, allowing organizations to be more consistent in how they apply deidentification in practice. Further, HHS could issue more specific guidance, or the health care community could form its own consortium or alliance to describe best practices or provide case studies on deidentification (McGraw 2013).

An example of this approach can be found in Europe, where the General Data Protection Regulation (GDPR) provides for an anonymization exemption to data processing that is very similar to deidentification under HIPAA (although it does explicitly say that pseudonymization is not the same as anonymization and is not exempted from GDPR oversight). While GDPR does not set an acceptable threshold of risk for reidentification, the European Medicines Agency (EMA), which is in charge of the evaluation and supervision of medicinal products in the European Union, sets its own standard. EMA's standard of a reidentification risk that does not exceed 9 percent aligns with the approach taken by the Centers for Medicare & Medicaid Services, which permits data to be published as a table only when each data cell represents 11 or more people. These standards demonstrate that it is possible to define the risk of reidentification more clearly within the context of how the data are used.

Redefining Safe Harbor

The safe harbor provision was created before the revolution in big data, when most identifiable information was found in insurance billing systems. At the time, collecting, sharing, and using health data occurred on a relatively small scale. The list of 18 categories for data exclusion was based on explicit identifiers—like names, which would lead directly to an individual's identification—and quasi-identifiers, like birth dates and ZIP codes, which could readily be linked to other sources to determine an individual's identity.

1. Name Last
First
2. Date of Birth (mm/dd/yyyy)
3. Sex
M F
4. Place of
5. Social Security Number
6. Email Address (e.g.
dress: Line 1: Street/RFD#, P.O. Box, or URB
learly label Apartment, Company, Suite, Unit

Over the past two decades, health data have become much more complex and varied. Datasets can exclude the 18 categories of data, but still contain all sorts of identifying information such as marriage status, sexual orientation, number of children in first marriage, number of children in second marriage, income level, type of occupation, and more. With the proliferation of data elements, some data that meet the safe harbor standards can nonetheless be reidentified.

It is time to shift from the current safe harbor provision to a robust deidentification approach. It is more appropriate to define the categories of data that can safely be included in deidentified data rather than to base the safe harbor on categories that must be excluded.

Harmonize Disparate Sectors

The United States has taken a sectoral approach here, with different laws regulating the protection of data from different domains. HIPAA applies only to personally identifiable health information, while state laws for consumer data protection extend more broadly to transactional relationships with businesses. As well, there are inconsistencies within health care itself. HIPAA affords privacy rights to individuals for 50 years after their death, while the Common Rule, which provides oversight on human subjects research, affords rights only to living individuals. This creates the possibility that research data from an EHR can be reidentified upon a person's death but should have been protected for much longer. Although the Common Rule does not use the term “deidentification,” it is implied by the way research participants are effectively severed from their data.

Unlike the United States, other regions of the world have created cross-sectoral laws and regulations. GDPR serves as an example where terms such as “personal data” and “identifiable natural person” have a uniform definition that is applied on a broad basis. Recent state-level consumer protection laws in the United States have begun to move in this direction, but they allow for exemptions for data generated in a HIPAA-covered environment. To ensure harmonization, it will be critical to remove such exemptions or allow industries to set consistent standards.

Bradley Malin, Ph.D., is the founder and co-director of the Health Data Science (HEADS) Center, the vice chair for research affairs in the Department of Biomedical Informatics at Vanderbilt University Medical Center, and the Accenture Professor of Biomedical Informatics, Biostatistics, and Computer Science at Vanderbilt University. He also co-directs the Center for Genetic Privacy and Identity in Community Settings (an NIH Center of Excellence on Ethical, Legal, and Social Implications Research), and the Infrastructure Core of the NIH Artificial

Intelligence/Machine Learning Consortium to Advance Health Equity and Researcher Diversity. Malin currently serves as the co-chair of the Committee on Access, Privacy, and Security of the All of Us Research Program (part of the U.S. Precision Medicine Initiative) and is an appointed member of both the Technical Anonymisation Group of the European Medicines Agency and the Board of Scientific Counselors of the CDC's National Center for Health Statistics. He is an elected fellow of the National Academy of Medicine, the American College of Medical Informatics, the International Academy of Health Sciences Informatics, and the American Institute for Medical and Biological Engineering.

References

Anastasijevic, D. (2020, September 23). Mayo Clinic completes deidentification of expansive medical dataset. Mayo Clinic News Network. https://newsnetwork.mayoclinic.org/discussion/mayo-clinic-completes-deidentification-of-expansive-medical-dataset/

Caplan, A., & Batra, P. (2019, November 14). The ethics of using deidentified medical data for research without informed consent. *Voices in Bioethics,* 5. https://doi.org/10.7916/vib.v5i.5917

Chen, M., Tan, X., & Padman, R. (2020). Social determinants of health in electronic health records and their impact on analysis and risk prediction: A systematic review. *Journal of the American Medical Informatics Association*, 27(11), 1764–1773. doi: 10.1093/jamia/ocaa143

Colicchio, T. K., Cimino, J. J., & Del Fiol, G. (2019). Unintended consequences of nationwide electronic health record adoption: Challenges and opportunities in the post-meaningful use era. *Journal of Medical Internet Research*, *21*(6), e13313. https://www.jmir.org/2019/6/e13313/

Corrigan-Curay, J., Sacks, L., and Woodcock, J. (2018). Real-world evidence and real-world data for evaluating drug safety and effectiveness. *Journal of the American Medical Association*, 320(9), 867–868. https://jamanetwork.com/journals/jama/article-abstract/2697359

Danciu, I., Cowan, J. D., Basford, M., Wang, X., Saip, A., Osgood, S., Shirey-Rice, J., Kirby, J., & Harris, P. A. (2014, December). Secondary use of clinical data: The Vanderbilt approach. *Journal of Biomedical Informatics*, 52, 28–35.

Donkin, L., Hickie, I. B., Christensen, H., Naismith, S. L., Neal, B., Cockayne, N. L., & Glozier, N. (2012). Sampling bias in an internet treatment trial for depression. *Translational Psychiatry*, 2(10), e174. https://doi.org/10.1038/tp.2012.100

El Emam, K., Jonker, E., Moher, E., & Arbuckle, L. (2013). A review of evidence on consent bias in research. *American Journal of Bioethics*, 13(4), 42–44. DOI: 10.1080/15265161.2013.767958

eMERGE Consortium. (2021). Lessons learned from the eMERGE Network: Balancing genomics in discovery and practice. *HGG Advances,* 2(1): 100018. https://www.sciencedirect.com/science/article/pii/S266624772030018X

Haendel, M. A., Chute, C. G., Bennett, T. D., Eichmann, D. A., Guinney, J., Kibbe, W. A., Payne, P. R. O., Pfaff, E. R., Robinson, P. N., Saltz, J. H., Spratt, H., Suver, C., Wilbanks, J., Wilcox, A. B., Williams, A. E., Wu, C., Blacketer, C., Bradford, R. L., Cimino, J. J., … N3C Consortium. (2021). The National COVID Cohort Collaborative (N3C): Rationale, design, infrastructure, and deployment. *Journal of the American Medical Informatics Association, 28*(3), 427–443. https://pubmed.ncbi.nlm.nih.gov/32805036/

HealthVerity. (2022). Discover RWD in a brand new way. https://healthverity.com/solutions/healthverity-marketplace/

Hripcsak, G., Mirhaji, P., Low, A. F. H., & Malin, B. A. (2016). Preserving temporal relations in clinical data while maintaining privacy. *Journal of the American Medical Informatics Association, 23*(6), 1040–1045. https://academic.oup.com/jamia/article/23/6/1040/2399243?login=false

IBM. (2022). Data, tools, and services designed for life sciences. https://www.ibm.com/products/marketscan-research-databases

Institute of Medicine. (2015, January 8). Committee on the Recommended Social and Behavioral Domains and Measures for Electronic Health Records; Board on Population Health and Health Practice. Capturing social and behavioral domains and measures in electronic health records: phase 2. National Academies Press.

Johnson, A., Pollard, T., & Mark, R. (2016, September 4). MIMIC-III clinical database. Physionet. https://physionet.org/content/mimiciii/1.4/

Kho, A. N., Cashy, J. P., Jackson, K. L., Pah, A. R., Goel, S., Boehnke, J., Humphries, J. E., Kominers, S. D., Hota, B. N., Sims, S. A., Malin, B. A., French, D. D., Walunas, T. L., Meltzer, D. O., Kaleba, E. O., Jones, R. C., & Galanter, W. L. (2015, September). Design and implementation of a privacy preserving electronic health record linkage tool in Chicago. *Journal of the American Medical Informatics Association, 22*(5), 1072–1080. https://academic.oup.com/jamia/article/22/5/1072/930113?login=false

Lee, B., Dupervil, B., Deputy, N. P., Duck, W., Soroka, S., Bottichio, L., Silk, B., Price, J., Sweeney, P., Fuld, F., Weber, J. T., & Pollock, D. (2021, September–October). Protecting privacy and transforming COVID-19 case surveillance datasets for public use. *Public Health Reports, 136*(5), 554–561. https://pubmed.ncbi.nlm.nih.gov/34139910/

Malin, B., Benitez, K., & Masys, D. (2011). Never too old for anonymity: A statistical standard for demographic data sharing via the HIPAA Privacy Rule. *Journal of the American Medical Informatics Association, 18*(1), 3–10. https://academic.oup.com/jamia/article/18/1/3/2909114?login=false

May, T. (2021). Introducing Datavant switchboard. Datavant. https://datavant.com/resources/blog/introducing-the-datavant-switchboard/

McGraw, D. (2013). Building public trust in uses of Health Insurance Portability and Accountability Act deidentified data. *Journal of the American Medical Informatics Association, 20*(1), 29–34. https://academic.oup.com/jamia/article/20/1/29/2909277

National Human Genome Research Institute. (2021). The cost of sequencing a human genome. National Institutes of Health. https://www.genome.gov/about-genomics/fact-sheets/Sequencing-Human-Genome-cost

National Institutes of Health. (2014). NIH genomic data sharing policy. NOT-0D-14-124. https://grants.nih.gov/grants/guide/notice-files/NOT-OD-14-124.html

Office for Civil Rights. (2013). Guidance regarding methods for deidentification of protected health information in accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule. U.S. Department of Health and Human Services. https://www.hhs.gov/guidance/document/guidance-regarding-methods-de-identification-protected-health-information-accordance

Office of Management and Budget. (2013). Supplemental guidance on the implementation of M-13-13: "Open data policy—Managing information as an asset." https://resources.data.gov/resources/m-13-13-guidance/

Perkel, J. M. (2019). The microscope makers putting ever-larger biological samples under the spotlight. *Nature*, 575 (7784): 715–717. https://www.nature.com/articles/d41586-019-03632-y

Roden, D. M., Pulley, J. M., Basford, M. A., Bernard, G. R., Clayton, E. W., Balser, J. R., & Masys, D. R. (2008, May 21). Development of a large-scale deidentified DNA biobank to enable personalized medicine. *Clinical Pharmacology and Therapeutics, 84*(3), 362–369. https://doi.org/10.1038/clpt.2008.89

Seykora, A., Coleman, C., Rosenfeld, S. J., Bierer, B. E., & Lynch, H. F. (2021). Steps toward a system of IRB precedent: Piloting approaches to summarizing IRB decisions for future use. *Ethics and Human Research, 43*(6), 2–18. https://onlinelibrary.wiley.com/doi/10.1002/eahr.500106

United States. (1996). Health Insurance Portability and Accountability Act (HIPAA) of 1996. Pub. L. 104-191, 110 Stat. 1936. https://www.govinfo.gov/content/pkg/PLAW-104publ191/pdf/PLAW-104publ191.pdf

Wei, W. Q., & Denny, J. C. (2015, April 30). Extracting research-quality phenotypes from electronic health records to support precision medicine. *Genome Medicine, 7*(41). https://doi.org/10.1186/s13073-015-0166-y

Weiskopf, N. G., Bakken, S., Hripcsak, G., & Weng, C. (2017). A data quality assessment guideline for electronic health data reuse. *EGEMS,* 5(1), 14. doi: 10.5334/egems.218

Wendler, D., Kington, R., Madans, J., Van Wye, G., Christ-Schmidt, H., Pratt, L. A., Brawley, O. W., Gross, C. P., & Emanuel, E. (2005, December). Are racial and ethnic minorities less willing to participate in health research? *PLoS Medicine,* 3(2), e19. doi: 10.1371/journal.pmed.0030019

Williams, R., Kontopantelis, E., Buchan, I., & Peek, N. (2017). Clinical code set engineering for reusing EHR data for research: A review. *Journal of Biomedical Informatics*, 70, 1–13. https://doi.org/10.1016/j.jbi.2017.04.010

"New investments and agreement on a robust data infrastructure that supports data standardization and interoperability are required if we are to achieve desired changes in health system design and practice."

– KENNETH D. MANDL, M.D., M.P.H.

The Value and Uses of Health Data in the Clinical Ecosystem

Kenneth D. Mandl, M.D., M.P.H.

Introduction

The processes and business of health and health care delivery increasingly rely on rapidly derived knowledge from large amounts of data on individual patients and populations. The last decade has seen remarkable gains in computing power, dramatically lower costs of data storage, the rise of a cloud computing industry, a jump to over 280 million smartphone users in the United States, plummeting costs in molecular measurements, and major advances in machine learning and artificial intelligence. The health care system has appropriately begun to turn its sights toward the use of these technologies for advancing individual care and population health.

The SARS-CoV-2 pandemic brought the health system's dependency on data and information processing into stark relief. Care facilities were overwhelmed with both patient volume and acuity. Public health struggled to understand the case definition and infection and fatality rates of COVID-19. Patients received mixed messages about the best actions to protect and treat themselves. Physicians and researchers attempted to design care regimens for severely ill patients infected with a novel virus.

In the face of overwhelming challenge there were important successes. To manage capacity and flow, hospital IT teams stood up dashboards tracking patient, staff, and community SARS-CoV-2 positivity rates and bed occupancy. Informaticians, epidemiologists, data journalists, and others made vast troves of data available to the public who became newly literate in key principles of infectious disease epidemiology—even down to the viral reproduction number.

A particularly bright spot during the early days of the pandemic was the Recovery Trial (RECOVERY Collaborative Group et al. 2021), a randomized trial embedded in the health care system. It has enrolled more than 47,000 participants at 200 hospital sites in six countries testing 10 therapies and validating four effective COVID-19 treatments (University of Oxford 2022). The study recruited patients with phenomenal speed, enrolling 12,000 subjects in its first 100 days. It established dexamethasone as an effective COVID-19 treatment for hospitalized patients on supplemental oxygen in a practice-changing preprint (Horby et al. 2020) published in June 2020, only three months after the study's initiation.

To evolve, the health system should learn from its experience, measure and improve value, modulate spending, better interface with the research enterprise and public health, and efficiently perform a vast array of knowledge management tasks, including ones as fundamental as diagnosis (Yang, Fineberg, & Cosby 2021). A movement in this direction has begun, building on a massive federal investment to promote the capture of encounter data through electronic health records (EHRs). There are emerging data exchange regimes, new technologies producing interoperable systems, novel governance models for intelligent data use across sites of care, and emerging business models for data aggregation.

Each of these advances has implications for patient autonomy, privacy, and protection from harm. Solving intractable problems in health demands a balance between protecting patients and making data available. Failure to protect patient privacy and shield individuals from unethical or harmful uses of data could upset this balance. Striking this balance is essential if we are to advance health care to more closely resemble the successes witnessed in the midst of the pandemic, rather than the chaotic information desert of its early days. This paper examines the infrastructure needs of a modern health care system and its relationship to the imperative of protecting health data privacy.

Data Needs for Health Care Improvement

New investments and agreement on a robust data infrastructure that supports data standardization and interoperability are required if we are to achieve desired changes in health system design and practice.

A Learning Health System

The RECOVERY trial is a leading example of the "learning health system" model (McGinnis, Fineberg, & Dzau 2021), whereby the processes around care and wellness continually produce data that yield insights. The National Academy of Medicine (Institute of Medicine (U.S.) Roundtable on Evidence-Based Medicine 2007) has long advocated for such a system, in which the outcomes of patients 1 through n are used to inform the treatment of patient n+1. A learning health system becomes smarter with every new patient, every encounter, and every action of the connected patient, whether at home, in school, or at the workplace.

This learning approach both complements and contrasts with the traditional model of evidence-based medicine, which leans heavily on information derived from traditional clinical trial results. The RECOVERY trial relies on curated routine datasets to track endpoints and outcomes (McCall 2021), effectively turning the delivery system into an engine for learning and research. Establishing a learning health system requires an intensive focus on data collection, curation, and analysis.

COVID-19's impacts on health care and public health (King 2020) demonstrate the necessity of health data liquidity (i.e., maximizing the proportion of health data that are easy to use and share) and standardization for any coordinated public health or health care improvement initiative. These advances are urgently needed throughout health care so that they can be applied broadly to standard practices of medicine and public health.

Supporting Value-Based Care

The health care sector is in the midst of a shift from fee-for-service payment models to payment tied to outcomes. Decoupling payment from individual services places a premium on measuring and achieving quality and value. Harnessing the data of the health system is central to this endeavor.

Risk-based contracts, whether capitation, shared savings, or bundled payments, include cost and quality metrics. However, measures of quality, such as the National Committee for Quality Assurance's Healthcare Effectiveness Data and Information Set (HEDIS) measures, are still typically communicated between payers and providers through periodic reporting, often using pdf reports or spreadsheets.

Measuring and improving value is a data-intensive proposition, and metrics should be generated in real time. They should be derived, at least in part, directly and consistently from data routinely collected in the health care system.

The first step in increasing data collection, availability, and use is to keep it in an electronic format. The American Recovery and Reinvestment Act of 2009 (ARRA, or Recovery Act), established the Health Information Technology for Economic and Clinical Health (HITECH) Act, through which the Centers for Medicare & Medicaid Services (CMS) disbursed more than $35 billion in incentive payments between 2011 and 2018 to more than 500,000 health care providers to adopt certified EHR systems (Centers for Medicare & Medicaid Services 2021).

CMS
CENTERS FOR MEDICARE & MEDICAID SERVICES

That federal investment was complemented by a much larger private investment by physicians and hospitals to purchase, install and configure EHR systems. These expenditures have resulted in over 96 percent of acute-care hospitals (Office of the National Coordinator for Health Information Technology 2022a) and 72 percent of office-based physicians using EHRs (Office of the National Coordinator for Health Information Technology 2022b).

The second step is using these electronic data to support real-time quality measurement and improvement. The Department of Health and Human Services (HHS) is advancing a program of electronic clinical quality measures (eCQM) that assess health care quality based on structured data collected during the process of patient care. Some measures draw data directly from the EHR, including clinician notes and laboratory test results, to examine care at a single care site. Others rely on claims data, which gather patient data across all sites of care, including medication information captured through the pharmacy benefit (Mandl & Mandel 2015). While most value-based payment methods examine quality measures only on an annual basis, the robust data behind eCQMs enable more refined analysis than is possible with legacy quality measures. Further, there is an opportunity to maximize the proportion of measures that are computable from available EHR data.

Prior Authorization

Prior authorization is the requirement that a patient (through their clinician) receives approval from the payer before receiving a service. Prior authorization is often employed for high-cost services or those that payers consider to be subject to overuse. Prior authorization requirements are a primary source of clinician frustration. Patients, caregivers, and clinicians face substantial information exchange barriers when seeking to justify insurance payment for an intervention or test. New electronic and automated prior authorization frameworks are being developed to reduce burden and redundancy, as well as to increase efficiency and improve care.

Reporting Efficiency

Clinicians, researchers, department chairs, and health system CEOs and CIOs face a mounting array of data access requests by myriad disparate governmental, scientific, and commercial constituencies. Hospitals already manage dozens or hundreds of mandated outbound data requirements from federal and state agencies, the Joint Commission, and numerous payers (Mandl & Kohane 2015). Each requires different formats that are resource intensive to create, and a medical practice or hospital must invest resources to respond accurately to these requests. Efficiencies of scale are within reach as many of the data types required under these regulatory regimes are core EHR elements and are also used to measure quality and value, creating an opportunity for convergence. An organization with access to a common set of EHR data elements can readily respond to myriad report types using scripted, programmable processes. Even greater efficiencies are possible when organizations limit their reporting requirements to a common set of data elements.

From a privacy perspective, an organization must decide what data it can legally and ethically report. Many reporting requirements involve only summary statistics, but others involve patient-level data. For example, when a public health authority requires data sharing under a mandatory reporting requirement for infectious diseases, the Health Insurance Portability and Accountability Act (HIPAA) of 1996 privacy rule does not apply. However, a hospital's decision to share patient data with a registry (such as one managed by a pharmaceutical company or patient advocacy group) requires a more complex calculus to determine whether the more appropriate approach is to obtain consent or a waiver of consent.

Improving Diagnostics

Physicians today make diagnoses much as William Osler did in the 19th century—by taking a history, examining the patient, and pattern matching to a traditional

taxonomy of medical conditions. This process depends upon clinical experience, considering symptoms, family history, and laboratory data, as well as interpretation of the medical literature. However, diagnosis, like other functions in medicine, is evolving to a data-driven science where algorithms learn from real-world examples, often derived from very large samples (Mandl & Bourgeois 2017). For example, in a landmark study, researchers from Google curated datasets with more than 11,000 retinal fundus images and derived a deep learning algorithm that outperformed clinicians for diagnosing diabetic retinopathy (Gulshan et al. 2016).

Much of modern clinical practice associated with interpreting genetic variations is based on studies of small numbers of individuals. The pitfalls of this approach were demonstrated when a reexamination of certain genetic variants, widely considered diagnostic of hypertrophic cardiomyopathy, were found to be normal among Black patients (Manrai et al. 2016). In another example, among 2,000 genotyped patients, the majority (41 of 63) with designated variants believed to be strongly associated with cardiac rhythm disturbances actually had no clinical evidence of the disorder (Van Driest et al. 2016).

Massive datasets must be assembled in order for artificial intelligence and genomics-informed medicine to yield the maximum knowledge derivable from the digitally captured experience of millions of patients. Making accurate diagnoses and determining effective treatment depend upon evaluating a patient's data against a large and representative population. The data needed to achieve this potential come not only from the EHR but must also link to biological samples, genomic sequence data, and the myriad data sources that capture medical care, behaviors, and environment.

To spare patients harm from diagnostic error, large-scale datasets are needed and it is in the interest of patients to understand that their health information contributes to these datasets (Committee on Diagnostic Error in Health Care et al. 2016) and will be used to benefit others. An expectation should be established that every patient benefits from diagnosis and treatment in a learning health care system. Toward this

goal, providers and their organizations should develop transparent and enabling compacts with their patients, perhaps expressed in privacy notices or in consent-to-treat documents (Mandl and Bourgeois 2017).

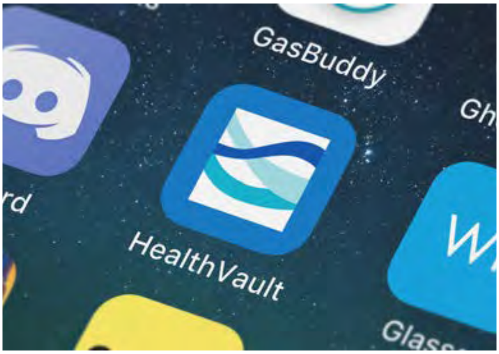
Infrastructure Needs

While much progress has been made in creating the robust infrastructure of data collection and sharing needed to gain the benefits of health data, more is required to realize the full potential of a learning health system. One critical barrier today is the inability to exchange and aggregate information across sites of care. Because the lion's share of health data is produced by provider organizations and stored locally, special solutions must emerge to produce a unified view of a patient or a population. Major issues include the willingness of organizations to share data, the readiness of health IT to export data, and the lack of common formats for data that are stored in health IT systems. Solutions may vary based on the use case being addressed. For example, the needs in patient care differ from those in pharmaceutical industry-led observational data analysis. Especially when the solutions involve third parties that are not covered entities under HIPAA, the regulatory framework for protecting patient privacy and preventing harm can be lacking.

Various models for collecting and aggregating data, and the privacy implications of those models, are discussed in this section.

Patient Control of Data

A patient-mediated model for exchanging health data is compelling. In theory, this model could address myriad use cases, such as defragmenting data for care or aggregating data for research. Since 1996, HIPAA has required health care organizations to provide patients with access to any data that are "readily producible," in the format the patient requests (Office of the National Coordinator for Health Information Technology 2015). For decades, organizations simply did not respond to patient requests, frequently citing HIPAA (incorrectly) as an excuse for refusing to transmit patient data. The patient's right of access to their own data was reasserted in HITECH, which requires organizations to provide the patient or their representative an electronic copy of their record (Miaoulis 2010). Yet patients continue to struggle to get copies of their records at all, much less in digital formats.



A sought-after alternative to patients seeking data from their providers is the personally controlled health record (PCHR) (Mandl & Kohane 2008; Mandl et al. 2007; Mandl, Szolovits, & Kohane 2001). PCHRs invert the current approach to medical records in that they reside with patients who then grant permission for their use to institutions, clinicians, researchers, and apps. The patient-control model aligns neatly with the HIPAA privacy rule requirements since the patient can always authorize sharing of their own data.

More than a decade ago, several PCHRs emerged based on an open-source model (Mandl et al. 2007), including the original Google Health, Microsoft HealthVault, and a claims-based PCHR for large employers called Dossia. None of these products was ultimately successful, in part because data export from EHRs was not available or standardized, making the consumer use case less compelling (Goldberg 2011). However, as discussed below, the ecosystem and its technologies are evolving rapidly, and new opportunities for patients to access health record data are emerging, including movements to enable patients to selectively share and monetize their data (Kostick-Quenet et al. 2022).

Policymakers are grappling with concerns relating to patients directing their data to third-party hosts. Unless the host is a HIPAA-covered entity, the data are regulated by the Federal Trade Commission (Sayeed et al. 2020) under Section 5(a) of the Federal Trade Commission Act, which places no explicit prohibitions on data reuse, only prohibiting "unfair or deceptive acts or practices in or affecting commerce" (Federal Trade Commission Act 2013). This raises the possibility that a predatory business could solicit patient data and be bound only by the terms and conditions agreed to, but not necessarily understood by, the patient. It is also unclear if and by whom breaches of those agreements would be enforced.

Large-Scale Commercial Aggregation

For the value-based care and 21st-century medicine use cases, EHR data must be combined across hospitals and clinics to create datasets reflecting large numbers of patients. Value-based care risk calculations require data on patients, potentially across many institutions in an accountable care organization, stratified by demographics and clinical characteristics. Rare disease research investigations need millions of patients to match study criteria; studies of gene variants, often with weak effects, require hundreds of thousands. Drug and device development depends on clinical trials that can require multiple sites to provide adequate statistical power. The 21st Century Cures Act of 2106 requires a program using real-world evidence—information, including EHR data, that is not derived from clinical trials—to support the approval of new indications for drugs and to satisfy post-approval study requirements.

Patient-mediated data exchange systems show promise but are not yet supported by a mature ecosystem. Instead, myriad commercial entities have arisen to enable bulk data exchange. Because of permissiveness under HIPAA, a common approach has been to rely on deidentified data. Sharing identified data between covered entities and their business associates, which is allowable for treatment, payment, and operations, has led to a torrent of EHR data flowing out of health care provider silos. HIPAA also allows business associates to deidentify data on behalf of the covered entity; when those data have been deidentified, the business associate may use them freely, if not contractually prohibited from doing so. Organizations that are not business associates under HIPAA may also receive and use deidentified datasets. This has enabled the rise of a multibillion-dollar industry comprising dozens of health data aggregation companies and hundreds of health data tool and technology companies aggregating, linking, and monetizing EHR data.

Some companies collect specific data on behalf of a customer. Ciox, for example, gathers medical records for a cohort by making record requests at all locations where members have received their care. Tokenization is an alternative approach—using cryptographic identifiers within two otherwise deidentified datasets to recognize the individual's identity and link that data. Datavant, for example, uses tokenization to help its customers aggregate patient-level data across multiple data sources. Datavant itself does not aggregate the data in the process.

Other business models include aggregating large reference datasets that can be used to derive intelligence—usually for care, discovery, product development, or marketing. Optum, IBM, and Truven sell access to massive research databases of patient data derived from payer claims and EHRs. Companies like IQVIA (which performs contract research), Medidata (an electronic data capture company), and Epic and Cerner (the dominant EHR vendors) contract with their users for data rights. The data are then licensed for use by Epic staff and researchers at sites contributing data. These arrangements are complex and take advantage of or sometime exceed the minimal regulatory framework around these data, as recently detailed in an investigation of the process across multiple companies of creating a massive, highly linked dataset involving tens of millions of Americans (Ross 2022).

Research Data Networks

Since the National Academy of Medicine first proposed a learning health system, many efforts to build the necessary data resources have encountered IT and structural challenges. For example, even obtaining and using data on a well-defined cohort in a standardized format at a single site is a challenge. Doing this across sites is harder still. Assembling a cohort large enough to study rare diseases, genomic variants, or the comparative effectiveness of drugs, or to match eligible patients to clinical trials, requires networks that reach across data systems and multiple sites.

A promising and often successful approach has been to aggregate data under a research regime, a tack potentially conducive to protecting patient privacy as the handling of identified research data falls under HIPAA regulations as overseen by

the HHS Office for Human Research Protections. The federal government has made enormous investments in multi-institutional efforts to align data in standardized ontologies across the Patient-Centered Clinical Research Network (Collins et al. 2014; Fleurence et al. 2014), National Institutes of Health (NIH) Accrual to Clinical Trials network (Visweswaran et al. 2018), and the NIH "All of Us" program (Denny et al. 2019). These investments support advances in comparative effectiveness measurements, clinical trials, and genomic medicine. A limitation of this approach is that obtaining and mapping the data require highly specialized IT personnel at any participating site, which can limit participation to more advanced hospitals, potentially leading to inequitable enrollment of populations (Bibbins-Domingo, Helman, & Dzau 2022).

Data Sharing Companies and Consortia

In order to practice 21st-century medicine, health systems must combine their own data with those from other systems. Federated networks (Mandl & Kohane 2015), like the Harvard SHRINE network (McMurry et al. 2013) and the Genomic Information Commons (Genomic Information Commons n.d.)—formerly the Genomic Research and Innovation Network (Mandl et al. 2020)—promote data sharing through self-governance and keep the data in place at each participating site. Other networks centralize the data (Forrest et al. 2014; Leverage Clinical and Resource Utilization Data n.d.).

23andMe
Personal Genome Service
Find out what your DNA says about you through online genetic reports.

In addition to hospital-led consortia, a new class of companies is contracting with health systems to help them aggregate and monetize their data. For example, TriNetX has developed a real-world data network where participating hospitals and health systems make their full EHR datasets available to be queried and accessed. Each site can receive deidentified data from other participating sites. In addition, TriNetX commercializes and sells the data without further participation of or authorization by the sites. Another example is Truvetta, led by a former Microsoft executive and founded and governed by health systems for the purpose of aggregating and making their data available to promote innovation among the member organizations and to commercialize and sell that data.

There are also myriad companies aggregating genomic data, including United States–based companies Invitae and Genuity and the direct-to-consumer company 23andMe, as well as international entities, such as China's Beijing Genomics Institute.

Health Information Exchange

A different set of privacy considerations and business models supports the exchange of data for care. In the last few years, the federal government has attempted to shore up systems that defragment a patient's record to provide a physician as accurate an understanding of medical history and interventions as possible. For example, at the beginning of 2021, the HHS Office of Civil Rights published proposed modifications to reduce barriers to coordinated care by, among other requirements, defining individual-level care coordination as a health care operation (Office for Civil Rights, Office of the Secretary, HHS 2021).

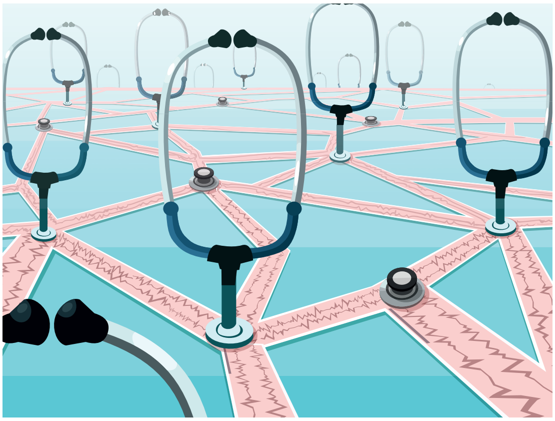
Because medical record data are primarily stored where they are produced, data become fragmented when patients traverse different sites of care. Historically, this problem has often been addressed by forming business entities that support data exchange. In the 1990s, there were some successful local models, such as the Indianapolis Network for Patient Care and Research (Overhage, Tierney, & McDonald 1995). Then a class of organizations, the Community Health Information Networks, emerged (Dullabh et al. 2011). In 2004, when the Office of the National Coordinator for Health Technology (ONC) was formed, the focus was on regional health information organizations, but after it became clear that not all central organizations supporting exchanging health information were regional, the more general term "health information exchange" (HIE) emerged.

ONC

Office of the National Coordinator for Health Information Technology

Many of the organizations, some supported by funds from HITECH, were not financially sustainable. For-profit HIEs emerged, such as Medicity (acquired by Aetna and then sold to Health Catalyst), as well as nonprofit organizations such as Manifest

Medex in California. An alternative model that has persisted is provider-to-provider information exchange, as supported by the federal Direct Project (Voigt & Torzewski 2011).



As specified in the 21st Century Cures Act, ONC has established a Trusted Exchange Framework and Common Agreement (TEFCA) to enable exchange across disparate networks (Office of the National Coordinator for Health Information Technology 2022c). The framework is a set of six principles articulating how data exchange should be conducted; the agreement establishes the legal and technical requirements for exchanging data over the network. Under TEFCA, qualified health information networks (QHINs) serve as exchange hubs, allowing participants, participant members, and individual users to request that electronic health information be sent or received through the QHIN Exchange Network.

Applications for health information networks are currently open though the ONC-supported nonprofit Sequoia Project. Careequality, launched by the Sequoia Project, is a nonprofit organization supporting nationwide information exchange among its members, including HINs, EHR vendors, personal health record vendors, and consumer groups.

The Epic ecosystem provides robust options for its customers. Care Everywhere enables Epic customers to share a subset of clinical data. A consortium of other EHR vendors and service providers, called the Commonwell Health Alliance, supports an HIE for non-Epic customers. Nonetheless, the HIE ecosystem still has important gaps.

Challenges to the Data Enterprise

There are challenges to achieving a learning health care system that relate to how data are acquired, stored, standardized, and shared. At every step of the process for data acquisition, sharing, analysis, and use, a robust privacy framework is essential to maintain patient and public trust.

Physician Burden

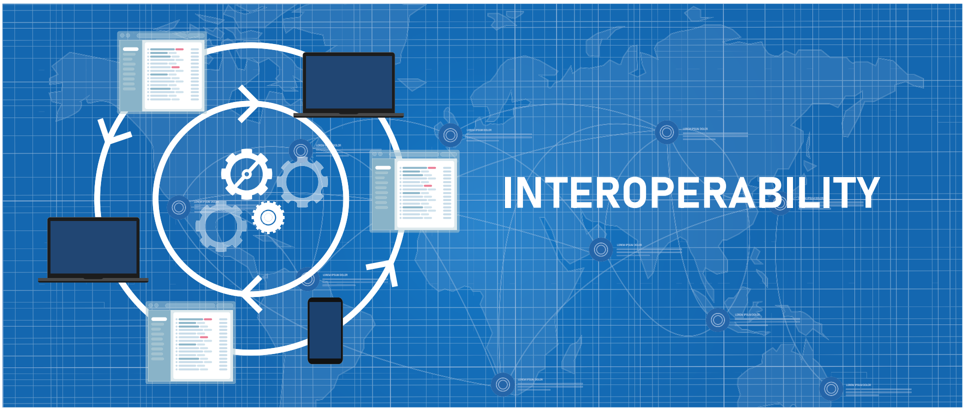
EHR technologies and their ecosystem need to evolve further to support the needs of care delivery. Most EHRs in current use were initially purchased primarily to manage the revenue cycle. They were not designed as modular systems, nor were they initially adept as tools to capture, exchange, and analyze data. Data production to support the business processes of health care is burdensome for physicians and can contribute to burnout (Kroth et al. 2019), as has been recognized by professional organizations (AMIA 25x5 2021; American Medical Association 2021), and a report of the Surgeon General (Murthy 2022).

An important factor is the extraordinary documentation requirement by payers to justify reimbursement (Downing, Bates, & Longhurst 2018). Another is that EHR workflows and user interfaces still need to evolve based on human-centered design (Shanafelt et al. 2016; Melnick et al. 2020). Addressing the problem will require workflow redesign, development of new methods in implementation science, and modernized approaches to software.

Lack of Standardization and Interoperability

EHRs store data in disparate formats, sometimes even across different installations of the same vendor product. Despite the vast amount of data that EHR systems collect, lack of interoperability across different vendor systems has limited how medical record data can be used to advance care and improve value (Mandl & Kohane 2019). Beyond technical interoperability, the cultures and values of the delivery and public health systems have often precluded voluntary and efficient sharing of data for care, discovery, and population health.

Interoperability should be a central goal for EHR software and data sharing among all stakeholders in order to construct business processes around them. Once a software application is developed, it should be able to connect and run anywhere in the health care system without one-off integrations. Once a statistical or analytic program is scripted, it should be able to run on the data produced by any EHR.



EHRs should be able to exchange data with one another so that a patient can have a complete medical record across sites of care and so that large datasets can be aggregated for analytic functions.

As discussed below, provisions of the 21st Century Cures Act and ensuing regulations begin to address standardization, interoperability, and practices around data exchange and sharing.

Data Privacy

Privacy is a fundamental right that protects individuals against abuses of power and supports their self-determination and individual preferences. It enables individuals to preserve their reputations, avoid stigma, and maintain their insurability. Yet even as the amount of data being produced and shared grows, and commercial interests in those data proliferate, the United States does not have a comprehensive data privacy law.

While patients frequently interact with HIPAA provisions, the public is often surprised to learn about the broad data uses allowed under HIPAA. In one case, Google acquired data under contract from two health systems to derive artificial intelligence algorithms under research protocols. Those contracts, even ones ultimately deemed HIPAA compliant and appropriate (Robbins 2020b), eventually drew scrutiny and some resulted in lawsuits (Robbins 2020a).

None of the various privacy-related laws or regulations protects patients from the potentially harmful use of deidentified data (Mandl & Perakslis 2021). The deidentification process is not infallible, and sometimes individuals can be reidentified on the basis of only a handful of attributes (Rocher, Hendrickx, & de Montjoye 2019).

Current encryption-based methods may be vulnerable to advances in computation. One approach to protecting individuals is to strengthen regulatory regimes—for example, around the use of deidentified data. States and the federal government could also pass laws against reidentifying deidentified health data, as California did.

In another example, under the Genetic Information Nondiscrimination Act, patients have special protections, particularly around use of their genetic data by employers and payers. The law also extends HIPAA privacy protections to genetic data, but notably applies only to covered entities and not to genomic companies, including direct-to-consumer genetic services; that is a gap that can be filled. Another approach is to establish best practices for data protection among data providers. Some efforts in this regard are underway. To facilitate health system participation, the company Health Evolution has convened a "Work Group on Governance and Use of Patient Data in Health IT Products" and is designing a trust framework for the data sharing used to develop algorithms. Another company, discussed below, the nonprofit Graphite Health, has published a "Digital Hippocratic Oath" as a pledge of trustworthy behavior when using and commercializing data from participating health systems (Graphite Health Vision 2022). While these efforts are well intentioned, their acceptability to health care providers has yet to be tested.

A Path to Interoperability

21st Century Cures Act

A federal rule goes into effect at the end of 2022 that addresses interoperability head on (Office of the National Coordinator for Health Information Technology 2020). The rule implements a provision of the 21st Century Cures Act of 2016 requiring that an application programming interface (API) be present in all certified health information technology. The API must afford "access to all data elements of a patient's EHR to the extent permissible under applicable privacy laws" with "no special effort" (United States 2016). The rule also prohibits information blocking, broadly defined as any practice likely to interfere with access, exchange, or use of electronic health information otherwise permitted by law. This new rule has tremendous implications for innovation and privacy, which are best understood by reviewing the provision's origins.

In 2009, as the Obama administration assumed office and proposed the HITECH Act, a public API to EHRs was proposed in the *New England Journal of Medicine* (Mandl & Kohane 2009). The iPhone was little more than a year old, and the API was central to enabling innovators to create applications (apps) that drew data from external sources into the phone (or another platform). Regardless of the platform of the original data, such apps could bring genomic information, machine learning–derived insights, and rich data visualizations to physicians in practice or directly to patients. The API would enable a reimagining of EHRs as smartphone-like platforms to which apps could be added or deleted. Apps include, for example, dynamically created, individualized, user-friendly medication instructions (Mandl, Gottlieb, & Ellis 2019) for patients available in many languages. An example for physicians is an app that helps manage blood pressure in children with hypertension, since blood pressure norms vary widely by age (Twichell et al. 2017).

A year later, the federal government funded the research and development of SMART (Substitutable Medical Applications, Reusable Technologies) (Computational Health Informatics Program n.d.). This is often referred to as "SMART on FHIR" (Mandel et al. 2016; Mandl et al. 2012) as it uses Fast Healthcare Interoperability Resources (FHIR) a standard for representing medical data developed by Health Level Seven (HL7). SMART on FHIR specifies a universal API so that patients or physicians can connect apps to EHRs. The SMART on FHIR API is also used for Medicare and Medicaid beneficiaries to obtain a copy of their claims data via the "Blue Button" (CMS Blue Button 2.0 n.d.).

Development of APIs can underpin a robust market for health apps (Mandl, Mandel, & Kohane 2015), foster innovation by individual entrepreneurs and giant technology companies, and vastly improve the health information technology experience for patients and their doctors (Kawamoto et al. 2021; Kawamoto et al. 2019).

APIs enable innovators to quickly gain scale, rather than face extensive delays while attempting to integrate their application with multiple EHRs. Over the past decade, companies creating SMART apps have been able to market and distribute their products through app stores and galleries (Mandl, Gottlieb, & Ellis 2019). SMART on FHIR apps compete with one another on features and value.

Apple has been the largest-scale user of the SMART on FHIR interface, connecting its native Health app to more than 800 health systems comprising more than 12,000 facilities and enabling 200 million iPhone users to acquire copies of their EHR in a computable format on their phones. Contributions by Epic, Cerner, and other EHR vendors, which implemented the API in their products, made this success possible. As a result, patients can share their data with providers, caretakers, an EHR, or an app. Notably, Apple has taken a strongly privacy-first stance by storing patient data only on the phone or encrypted in the user's iCloud backup. Apple cannot read or use the health data unless explicitly permitted by the patient.

A second SMART API is the Bulk FHIR Access API (Mandl et al. 2020), which draws population data into a uniform, standardized, computable format. Making the data requests turnkey and standardizing the output create a major opportunity for scalable analytic processes that can be created once and run anywhere in the health care system.

The regulatory framework is no different for data accessed using the Bulk FHIR API. A HIPAA-covered entity will exchange its FHIR data under normal HIPAA requirements, with less burden on the organization to perform manual extraction, transformation, and loading processes. The data extracted from a provider organization are protected under the same privacy and security laws and are only shared with a payer within the bounds of the existing contract, and vice versa.

The 21st Century Cures Act rule (Office of the National Coordinator for Health Information Technology 2020) specifically requires the SMART on FHIR and Bulk FHIR APIs to be supported in all health IT by the end of 2022.

Emerging Ecosystem

New organizations are emerging to take advantage of regulated interoperability. A growing class of organizations, known as FHIR accelerators, drive specific use cases for FHIR. Argonaut was the first of these, initially convening HL7, the SMART Team, and several EHR vendors to implement SMART on FHIR in the EHR products to meet a regulatory API requirement. Commure and Smile CDR (for-profit entities) and Graphite Health (a nonprofit) partner with health care institutions to provide a middleware layer enabling a SMART on FHIR-based app ecosystem for providers. Health care institutions hire Redox and Apigee to implement API and other functionality on top of EHRs. The large cloud vendors (Google Cloud, Microsoft, AWS, Oracle, IBM) have all built SMART APIs into their products.

Conclusion

The extant data-sharing regimes defined by patient or business associate agreements do not always prioritize patient privacy, autonomy, and insurability, even as they strive to maximize data use for a learning health care system. Expectations of increasingly brisk data flows when the interoperability provision of the 21st Century

Cures Act takes effect at the end of 2022 highlight the outdated patient protections of HIPAA, which did not anticipate electronic data sharing at scale.

There are many challenges to anticipate. Even now, patients usually don't know that HIPAA permits their identified information to be shared with any business associate of the covered entities providing their care for purposes of treatment, payment, and operations.

If the goal is a learning health system with standardized datasets yielding improved care and insights into disease causes and treatments, it is incumbent upon us to monitor the ecosystem, continually refining regulatory and legal frameworks and behavioral expectations for health systems, third-party apps, and companies aggregating and commercializing data. By doing so, we can ensure that the patient's perspective continues to hold primacy (Mandl & Kohane 2020) and that the health system learns, becomes smarter, improves efficiency, and becomes more affordable.

Kenneth D. Mandl, M.D., M.P.H., directs the Computational Health Informatics Program at Boston Children's Hospital and is the Donald A. B. Lindberg Professor of Pediatrics and Biomedical Informatics at Harvard Medical School. A real-time biosurveillance pioneer, he developed "SMART on FHIR," which allows patients and doctors to access an "app store for health." Through his influence on the 21st Century Cures Act, federal regulations require support for SMART interfaces, readily ensuring standardized access to individual and population data at system scale. Mandl also leads the federated Genomic Information Commons across nine children's hospitals. An elected member of the National Academy of Medicine, American Society for Clinical Investigation, Society for Pediatric Research, American College of Medical Informatics, and American Pediatric Society, Mandl has served as advisor to two CDC directors and board chair of Scientific Counselors of the NIH's National Library of Medicine.

References

American Medical Association. (2021. April 30). 5 simple changes to help cut doctors' EHR burdens. https://www.ama-assn.org/practice-management/digital/5-simple-changes-help-cut-doctors-ehr-burdens

AMIA 25x5. (2021). AMIA—American Medical Informatics Association. https://amia.org/about-amia/amia-25x5

Bibbins-Domingo, K., Helman, A., & Dzau, V. J. (2022, May 17). The imperative for diversity and inclusion in clinical trials and health research participation. *Journal of the American Medical Association, 327*(23), 2283–2284. https://doi.org/10.1001/jama.2022.9083

Centers for Medicare & Medicaid Services. (2021, December 1). Regulations & guidance—promoting interoperability: Data and program reports. https://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/DataAndReports

CMS Blue Button 2.0. (n.d.). Retrieved June 1, 2022, from https://bluebutton.cms.gov/

Collins, F. S., Hudson, K. L., Briggs, J. P., & Lauer, M. S. (2014). PCORnet: Turning a dream into reality. *Journal of the American Medical Informatics Association, 21*(4), 576–577. doi: 10.1136/amiajnl-2014-002864

Committee on Diagnostic Error in Health Care, Board on Health Care Services, Institute of Medicine, and the National Academies of Sciences, Engineering, and Medicine. (2016). E. Balogh, B. Miller, and J. Ball (Eds.), *Improving diagnosis in health care*. National Academies Press.

Computational Health Informatics Program. (n.d.). SMART health IT. Retrieved May 30, 2022, from https://smarthealthit.org/

Denny, J. C., Rutter, J. L., Goldstein, D. B., Philippakis, A., Smoller, J. W., Jenkins, G., & Dishman, E. (2019). The "All of Us" research program. *New England Journal of Medicine, 381*(7), 668–676. doi: 10.1056/NEJMsr1809937. PMID: 31412182; PMCID: PMC8291101

Downing, N. L., Bates, D. W., & Longhurst, C. A. (2018). Physician burnout in the electronic health record era: Are we ignoring the real cause? *Annals of Internal Medicine, 169*(1): 50–51. DOI: 10.7326/M18-0139

Dullabh, P., Moiduddin, A., Nye, C., & Virost, L. (2011). The evolution of the state health information exchange cooperative agreement program: State plans to enable robust HIE. HealthIT.gov. https://www.healthit.gov/sites/default/files/pdf/state-health-info-exchange-program-evolution.pdf

Federal Trade Commission Act. (2013, July 19). Federal Trade Commission. 15 U.S.C. §§ 41-58, as amended. https://www.ftc.gov/legal-library/browse/statutes/federal-trade-commission-act

Fleurence, R. L., Curtis, L. H., Califf, R. M., Platt, R., Selby, J. V., & Brown, J. S. (2014). Launching PCORnet, a national patient-centered clinical research network. *Journal of the American Medical Informatics Association*, 21(4), 578–582. DOI: 10.1136/amiajnl-2014-002747

Forrest, C. B., Margolis, P. A. Bailey, L. C., Marsolo, K., Del Beccaro, M. A., Finkelstein, J. A., Milov, D. E., Vieland, V. J., Wolf, B. A., Yu, F. B., & Kahn, M. G. (2014, May). PEDSnet: A national pediatric learning health system. *Journal of the American Medical Informatics Association,* 21(4), 602–606. doi: 10.1136/amiajnl-2014-002743

Genomic Information Commons. (n.d.). Retrieved June 1, 2022, from https://www.genomicinformationcommons.org/

Goldberg, C. (2011, July 27). What killed Google Health? And what does its untimely demise mean? *CommonHealth*.

Graphite Health Vision. (2022). GraphiteHealth.Io. https://www.graphitehealth.io/vision

Gulshan, V., Peng, L., Coram, M., Stumpe, M. C., Wu, D., Narayanaswamy, A., Venugopalan, S., Widner, K., Madams, T., Cuadros, J., Kim, R., Raman, R., Nelson, P. C., Mega, J. L., & Webster, D. R. (2016). Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *Journal of the American Medical Association, 316*(22): 2402–2410. doi: 10.1001/jama.2016.17216

Horby, P., Lim, W. S., Emberson, J., Mafham, M., Bell, J., Linsell, L., Staplin, N., Brightling, C., Ustianowski, A., Elmahi, E., Prudon, B., Green, C., Felton, T., Chadwick, D., Rege, K., Fegan, C., Chappell, L., Faust, S., Jaki, T., … RECOVERY Collaborative Group. (2020, June). Effect of dexamethasone in hospitalized patients with COVID-19—preliminary report. *MedRxiv*. https://doi.org/10.1101/2020.06.22.20137273

Institute of Medicine (U.S.) Roundtable on Evidence-Based Medicine. (2007). L. Olsen, D. Aisner, and J. McGinnis (Eds.). *The learning healthcare system: Workshop summary.* National Academies Press.

Kawamoto, K., Kukhareva, P., Shakib, J. H., Kramer, H., Rodriguez, S., Warner, P. B., Shields, D., Weir, C., Del Fiol, G., Taft, T., & Stipelman, C. H. (2019, November 1). Association of an electronic health record add-on app for neonatal bilirubin management with physician efficiency and care quality. *JAMA Network Open*, 2(11), e1915343. doi: 10.1001/jamanetworkopen.2019.15343

Kawamoto, K., Kukhareva, P. V., Weir, C., Flynn, M. C., Nanjo, C. J., Martin, D. K., Warner, P. B., Shields, D. E., Rodriguez-Loya, S., Bradshaw, R. L., Cornia, R. C., Reese, T. J., Kramer, H. S., Taft, T., Curran, R. L., Morgan, K. L., Borbolla, D., Hightower, M., Turnbull, W. J., … Del Fiol, G. (2021). Establishing a multidisciplinary initiative for interoperable electronic health record innovations at an academic medical center. *JAMIA Open, 4*(3), 1–15. doi: 10.1093/jamiaopen/ooab041

King, J. S. (2020). Covid-19 and the need for health care reform. *New England Journal of Medicine, 382*(26), e104.

Kostick-Quenet, K., Mandl, K. D., Minssen, D., Cohen, I. G., Gasser, U., Kohane, I., & McGuire, A. L. (2022). How NFTs could transform health information exchange. *Science,* 375(6580), 500–502. https://pubmed.ncbi.nlm.nih.gov/35113709/

Kroth, P. J., Morioka-Douglas, N., Veres, S., Babbott, S., Poplau, S., Qeadan, F., Parshall, C., Corrigan, K., & Linzer, M. (2019, August 2). Association of electronic health record design and use factors with clinician stress and burnout. *JAMA Network Open,* 2(8), e199609. doi: 10.1001/jamanetworkopen.2019.9609

Leverage Clinical and Resource Utilization Data. (n.d.). Children's Hospital Association. Retrieved June 8, 2022, from https://www.childrenshospitals.org/content/analytics/product-program/pediatric-health-information-system

Mandel, J. C., Kreda, D. A., Mandl, K. D., Kohane, I. S., & Ramoni, R. B. (2016). SMART on FHIR: A standards-based, interoperable apps platform for electronic health records. *Journal of the American Medical Informatics Association, 23*(5), 899–908. https://pubmed.ncbi.nlm.nih.gov/26911829/

Mandl, K. D., & Bourgeois, F. T. (2017). The evolution of patient diagnosis: From art to digital data-driven science. *Journal of the American Medical Association, 318*(19), 1859–1860.

Mandl, K. D., Glauser, T., Krantz, I. D., Avillach, P., Bartels, A., Beggs, A. H., Biswas, S., Bourgeois, F. T., Corsmo, J., Dauber, A., Devkota, B., Fleisher, G. R., Heath, A. P., Helbig, I., Hirschhorn, J. N., Kilbourn, J., Kong, S. W., Kornetsky, S., Majzoub, J. A., … Genomics Research and Innovation Network. (2020). The Genomics Research and Innovation

Network: Creating an interoperable, federated, genomics learning system. Genetics in Medicine: Official Journal of the American College of Medical Genetics, 22(2), 371–380. https://pubmed.ncbi.nlm.nih.gov/31481752/

Mandl, K. D., Gottlieb, D., & Ellis, A. (2019). Beyond one-off integrations: A commercial, substitutable, reusable, standards-based, electronic health record–connected app. *Journal of Medical Internet Research, 21*(2), e12902.

Mandl, K. D., Gottlieb, D., Mandel, J. C., Ignatov, V., Sayeed, R., Grieve, G., Jones, J., Ellis, A., & Culbertson, A. (2020). Push button population health: The SMART/HL7 FHIR bulk data access application programming interface. *npj Digital Medicine, 3*(1), 1–9. https://pubmed.ncbi.nlm.nih.gov/33299056/

Mandl, K. D., & Kohane, I. S. (2008). Tectonic shifts in the health information economy. *New England Journal of Medicine, 358*(16), 1732–1737.

Mandl, K. D., & Kohane, I. S. (2009). No small change for the health information economy. *New England Journal of Medicine, 360*(13), 1278–81.

Mandl, K. D., & Kohane, I. S. (2015). Federalist principles for healthcare data networks. *Nature Biotechnology, 33*(4), 360–363.

Mandl, K. D., & Kohane, I. S. (2019). Data standards may be wonky, but they will transform health care. StatNews. https://www.statnews.com/2019/10/03/data-standards-wonky-transform-health-care/

Mandl, K. D., & Kohane, I. S. (2020). Data citizenship under the 21st Century Cures Act. *New England Journal of Medicine, 382*(19), 1781–1783.

Mandl, K. D., & Mandel, J. C. (2015, April). Building a self-measuring healthcare system with computable metrics, data fusion, and substitutable apps. *BMJ Outcomes, 2015*(1), 6–13.

Mandl, K. D., Mandel, J. C., & Kohane, I. S. (2015). Driving innovation in health systems through an apps-based information economy. *Cell Systems, 1*(1), 8–13.

Mandl, K. D., Mandel, J. C., Murphy, S. N., Bernstam, E. V., Ramoni, R. L., Kreda, D. A., McCoy, J. M., Adida, B., & Kohane, I. S. (2012). The SMART platform: Early experience enabling substitutable applications for electronic health records. *Journal of the American Medical Informatics Association, 19*(4), 597–603. https://pubmed.ncbi.nlm.nih.gov/22427539/

Mandl, K. D., & Perakslis, E. D. (2021). HIPAA and the leak of "deidentified" EHR data. *New England Journal of Medicine, 384*(23), 2171–2173. DOI: 10.1056/NEJMp2102616

Mandl, K. D., Simons, W. W., Crawford, W. C. R., & Abbett, J. M. (2007, September 12). Indivo: A personally controlled health record for health information exchange and communication. *BMC Medical Informatics and Decision Making, 7,* 25. https://doi.org/10.1186/1472-6947-7-25

Mandl, K. D., Szolovits, P., & Kohane, I. S. (2001). Public standards and patients' control: How to keep electronic medical records accessible but private. *BMJ, 322*(7281), 283–287. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1119527/

Manrai, A. K., Funke, B. H., Rehm, H. L., Olesen, M., Maron, B. A., Szolovits, P., Margulies, D. M., Loscalzo, J., & Kohane, I. S. (2016). Genetic misdiagnoses and the potential for health disparities. *New England Journal of Medicine, 375*(7), 655–665. DOI: 10.1056/NEJMsa1507092

McCall, B. (2021, April). Data, data all around. *The Lancet*. Digital Health. https://doi.org/10.1016/S2589-7500(21)00063-7

McGinnis, J. M., Fineberg, H. V., & Dzau, V. J. (2021). Advancing the learning health system. *New England Journal of Medicine,* 385(1), 1–5. DOI: 10.1056/NEJMp2103872

McMurry, A. J., Murphy, S. N., MacFadden, D., Weber, G., Simons, W. W., Orechia, J., Bickel, J., Wattanasin, N., Gilbert, C., Trevvett, P, Churchill, S., & Kohane, I. S. (2013, March 7). SHRINE: Enabling nationally scalable multi-site disease studies. *PloS One, 8*(3), e55811. https://doi.org/10.1371/journal.pone.0055811

Melnick, E. R., Dyrbye, L. N., Sinsky, C. A., Trockel, M., West, C. P., Nedelec, L., Tutty, M. A., & Shanafelt, T. (2020). The association between perceived electronic health record usability and professional burnout among US physicians. *Mayo Clinic Proceedings,* 95(3), 476–487. DOI: 10.1016/j.mayocp.2019.09.024

Miaoulis, W. M. (2010). Access, use, and disclosure: HITECH's impact on the HIPAA touchstones. AHIMA. http://bok.ahima.org/doc?oid=98645

Murthy, V. H. (2022). Addressing health worker burnout: The U.S. surgeon general's advisory on building a thriving health workforce. U.S. Department of Health and Human Services. https://www.hhs.gov/sites/default/files/health-worker-wellbeing-advisory.pdf

Office for Civil Rights, Office of the Secretary, HHS. (2021). Proposed modifications to the HIPAA Privacy Rule to support, and remove barriers to, coordinated care and individual engagement. https://www.govinfo.gov/content/pkg/FR-2021-01-21/pdf/2020-27157.pdf

Office of the National Coordinator for Health Information Technology. (2015). Guide to privacy and security of electronic health information. https://www.healthit.gov/sites/default/files/pdf/privacy/privacy-and-security-guide.pdf

Office of the National Coordinator for Health Information Technology. (2020). ONC's Cures Act Final Rule. https://www.healthit.gov/curesrule/resources/information-blocking-faqs

Office of the National Coordinator for Health Information Technology. (2022a, April). Adoption of electronic health records by hospital service type, 2019–2021, Health IT Quick Stat #60. https://www.healthit.gov/data/quickstats/adoption-electronic-health-records-hospital-service-type-2019-2021

Office of the National Coordinator for Health Information Technology. (2022b, March). National trends in hospital and physician adoption of electronic health records. Health IT Quick-Stat #61. https://www.healthit.gov/data/quickstats/national-trends-hospital-and-physician-adoption-electronic-health-records

Office of the National Coordinator for Health Information Technology. (2022c). The Trusted Exchange Framework (TEF): Principles for trusted exchange. https://www.healthit.gov/sites/default/files/page/2022-01/Trusted_Exchange_Framework_0122.pdf

Overhage, J. M., Tierney, W. M., & McDonald, C. J. (1995). Design and implementation of the Indianapolis Network for Patient Care and Research. *Bulletin of the Medical Library Association, 83*(1), 48–56.

RECOVERY Collaborative Group, Horby, P., Lim, W. S., Emberson, J., Mafham, M., Bell, J. L., Linsell, L., Staplin, N., Brightling, C., Ustianowski, A., Elmahi, E., Prudon, B., Green, C., Felton, T., Chadwick, D., Rege, K., Fegan, C., Chappell, L. C., Faust, S. N., Jaki, T., … Landray, M. J. (2021, February 25). Dexamethasone in hospitalized patients with Covid-19. *New England Journal of Medicine, 384*(8): 693–704. doi: 10.1056/NEJMoa2021436

Robbins, R. (2020a, February 26). Contract offers unprecedented look at Google deal to obtain patient data from the University of California. STAT. https://www.statnews.com/2020/02/26/patient-data-contract-google-university-of-california/

Robbins, R. (2020b, March 3). Contract spells out how the University of Chicago shared patient data with Google. STAT. https://www.statnews.com/2020/03/03/contract-spells-out-how-the-university-of-chicago-shared-patient-data-with-google/

Rocher, L., Hendrickx, J. M., & de Montjoye, Y.-A. (2019). Estimating the success of re-identifications in incomplete datasets using generative models. *Nature Communications, 10*(3069). https://doi.org/10.1038/s41467-019-10933-3

Ross, C. (2022, May 23). How a complex web of businesses turned private health records from GE into a lucrative portrait of patients. STAT. https://www.statnews.com/2022/05/23/hipaa-patient-ge-data-privacy-profit/

Sayeed, R., Jones, J., Gottlieb, D., Mandel, J. C., & Mandl, K. D. (2020, December). A proposal for shoring up Federal Trade Commission protections for electronic health record–connected consumer apps under 21st Century Cures. *Journal of the American Medical Informatics Association, 28*(3). https://doi.org/10.1093/jamia/ocaa227

Shanafelt, T. D., Dyrbye, L. N., Sinsky, C., Hasan, O., Satele, D., Sloan, J., & West, C. P. (2016). Relationship between clerical burden and characteristics of the electronic environment with physician burnout and professional satisfaction. *Mayo Clinic Proceedings, 91*(7), 836–848.

Twichell, S. A., Rea, C. J., Melvin, P., Capraro, A. J., Mandel, J. C., Ferguson, M. A., Nigrin, D. J., Mandl, K. D., Graham, D., & Zachariah, J. P. (2017). The effect of an electronic health record–based tool on abnormal pediatric blood pressure recognition. *Congenital Heart Disease, 12*(4), 484–490.

United States. (2016). 21st Century Cures Act. Pub. L. No. 114-255. https://www.congress.gov/bill/114th-congress/house-bill/34/text

University of Oxford. (2022, March 24). The RECOVERY trial—two years on. https://www.ox.ac.uk/news/features/recovery-trial-two-years

Van Driest, S. L., Wells, Q. S., Stallings, S., Bush, W. S., Gordon, A., Nickerson, D. A., Kim, J. H., Crosslin, D. R., Jarvik, G. P., Carrell, D. S., Ralston, J. D., Larson, E. B., Bielinski, S. J., Olson, J. E., Ye, Z., Kullo, I. J., Abul-Husn, N. S., Scott, S. A., Bottinger, E., Almoguera, B., … Roden, D. M. (2016, January 5). Association of arrhythmia-related genetic variants with phenotypes documented in electronic medical records. *Journal of the American Medical Association, 315*(1), 47–57. doi 10.1001/jama.2015.17701

Visweswaran, S., Becich, M. J., D'Itri, V. S., Sendro, E. R., MacFadden, D., Anderson, N. R., Allen, K. A., Ranganathan, D., Murphy, S. N., Morrato, E. H., Pincus, H. A., Toto, R., Firestein, G. S., Nadler, L. M., & Reis, S. E. (2018, October). Accrual to Clinical Trials (ACT): A Clinical and Translational Science Award Consortium Network. *JAMIA Open, 1*(2), 147–152. doi: 10.1093/jamiaopen/ooy033

Voigt, C., & Torzewski, S. (2011). Direct results: An HIE tests simple information exchange using the Direct Project. *Journal of AHIMA / American Health Information Management Association, 82*(5), 38–41.

Yang, D., Fineberg, H. V., & Cosby, K. (2021). Diagnostic excellence. *Journal of the American Medical Association, 326*(19), 1905–1906. DOI: 10.1001/jama.2021.19493

Image Citations

Learning Health Systems
Practice
Evidence
Data
Research
Systematically gather and create evidence.
Apply the most promising evidence to improve care.

P.6: About Learning Health Systems. Content last reviewed May 2019. Agency for Healthcare Research and Quality, Rockville, MD. https://www.ahrq.gov/learning-health-systems/about.html

P.117: Daniel X. O'Neil https://www.flickr.com/photos/juggernautco/

Health Data Protection

PROTECTING HEALTH DATA PRIVACY *and* IMPROVING PATIENT CARE

A Report of the Aspen Health Strategy Group

The mission of the Aspen Health Strategy Group is to promote improvements in policy and practice by providing leadership on important and complex health issues. AHSG brings together senior leaders representing a mix of influential sectors, including health, business, philanthropy, and technology, to tackle a single health issue annually through year-long, in-depth study. Co-chairs are Kathleen Sebelius, 21st U.S. Secretary of Health and Human Services and former Governor of the State of Kansas, and William Frist, former U.S. Senator from Tennessee and former Senate Majority Leader.

The topic of AHSG's seventh annual report is protecting health data privacy and improving patient care. This compilation opens with a consensus report based on the group's in-depth learning process, followed by a set of background papers. Taken together, these papers explore the positive transformative power of health data, but also the many privacy challenges and concerns raised by "big data" collection, use, and analytics.

HEALTH
STRATEGY GROUP
aspen institute

HEALTH
MEDICINE
& SOCIETY
aspen institute

support for this report provided by:

Robert Wood Johnson
Foundation

Google Health

Laurie M Tisch
ILLUMINATION FUND