Health is a cornerstone of equitable societies. The 1948 Universal Declaration of Human Rights included the right to health, defining it as the highest attainable standard of physical and mental health. The broad definition of health as a right creates claims not just to medical treatment, but also to important determinants of health—from clean drinking water and adequate sanitation to political and economic equality. Individuals’ private health-related information is necessary to build and improve healthcare systems and public health and, therefore, to fulfill the right to health. As such, the right to health may be in tension with the right to privacy. This tension has been heightened during the Covid-19 pandemic and, indeed, sits at the center of debates regarding manual, digital, and automated contact tracing as well as public health surveillance for pandemic response more generally.

The historical trajectory of public health surveillance in the United States has been characterized by the increased collection and aggregation of health-related information. Local and state actors began to collect data to control the spread of infectious diseases like smallpox, cholera, typhoid, and tuberculosis in the late 1800s, expanding to information-gathering about a range of other diseases throughout the 20th century. Following the terrorist attacks of September 11, 2001, heightened concern over future attacks led to enactment of the US Public Health Security and Bioterrorism Preparedness and Response Act of 2002 and the subsequent establishment of the National Syndromic Surveillance Program (NSSP), consolidating the collection of near-real-time health information into a nationwide, cross-sector system. Under this program, individual and population health data, symptoms, laboratory findings, patterns of individual behavior (e.g., over-the-counter drug purchasing, school or workplace absenteeism owing to illness, etc.), and other information is collected from a variety of sources, including hospitals and clinics, pharmacies, places of employment, and primary and secondary schools. The syndromic surveillance system was developed initially for the early detection of bioterrorism and also public health emergencies; at full capacity, this infrastructure may yield data useful for public health, clinical medicine, and health research. However, in keeping with an expected pattern of diminishing fiscal support for programs put into place under emergency circumstances after the crisis has abated, the US Centers for Disease Control and Prevention (CDC) budget for the NSSP has been decreasing steadily in recent years, and by the time Covid-19 emerged only 70 percent of US hospitals were importing data into the national system.
Although there is not widespread public awareness of the existence of the NSSP and other public health surveillance measures, many patients are cognizant of the broad public benefit to healthcare research and medical knowledge that can result from the dissemination of their individual health information and may be willing to share it for this purpose. In the public health sector, positive reasons to share health information include the ability to disaggregate it in order to begin to understand variation in rates of infections, mortality, and morbidity within populations.

Yet sharing information also creates risks. Accordingly, some may be understandably reluctant to share personal information for medical purposes or health research. African Americans, for example, have been subject to a long and troubling history of encounters with unethical and exploitative research. This history has “affected their health-seeking behavior” and some African Americans are understandably distrustful of health and medical science and shy away from research participation. And while the collection and dissemination of health data can lead to health-improving, life-saving interventions, this practice may also make individuals and groups vulnerable to discrimination. The risks of discrimination against individuals are especially heightened as granular data, coupled with increasingly sophisticated computing techniques, make it difficult to judge whether people are truly not identifiable from their data. In addition, insurers, loan guarantors, employers, and others have financial incentives to exclude or otherwise discriminate against individuals based upon information about their health conditions. Health information may also subject groups to discrimination in instances in which risk factors are aligned with demographic categories, creating the possibility of collective stigma. Medical privacy therefore concerns more than a balance between the right to health and the right to privacy, or individual rights versus collective good; it also concerns group rights.

Below we review approaches to the regulation of health data in the US and the United Kingdom. While the US takes a sector-based approach to data privacy, the UK and European Union use a general framework that is adaptable to a range of sectors and contexts. The differences between these approaches meaningfully affects what data can be collected, by whom, and for what purpose. As Covid-19 motivates governments, corporate actors, and academics to collect more data for public health, it also demonstrates how many types of data—including information like location or social networks—are data about health and therefore health-relevant. The pace of contemporary scientific and technological development poses a challenge to more static health privacy regulation and ethical norms; the pandemic has exacerbated this dynamic.
When Is Data Health Data?

Public health authorities and healthcare systems in the US have made use of existing forms of data-based surveillance such as the NSSP and its state-level analogs. New forms of data are also being employed to help illuminate and address the pandemic crisis, sometimes without precedent or oversight. With the deployment of new kinds of health data come new regulatory challenges; among these is the question of what data is considered health data under existing norms and laws. As well as the question of how to regulate data that was not originally intended to be health-relevant data, but is being used for this purpose in the context of a public health or other emergency. Health data privacy regulation in the US offers an illustrative example of how health information has been defined in one context, and how that definition fails to include oversight about this data (or inferences that might be drawn from it) in other contexts.

The Health Insurance Portability and Accountability Act of 1996 (HIPAA) is the primary piece of national legislation that regulates health data in the US. It creates standards for proper storage, use, and transmission of personal health information handled by “covered entities” (clinicians, healthcare facilities, pharmacies, health plans, healthcare clearinghouses) in order to protect the confidentiality, integrity, and security of patient data. Simply put, HIPAA covers information created within most healthcare institutions. Healthcare providers may share information to the extent minimally necessary for treatment, payment, or healthcare operations. Patients must consent for providers to disclose individually identifiable information, but de-identified records may be used without consent for some purposes, like infectious disease control. Checks on this law include the use of federal auditors to detect privacy-violating or otherwise problematic data practices in healthcare institutions and the requirement that doctors’ offices must give patients records of their data if they request them.

There may be troubling downsides to the individual “data determination” afforded by HIPAA. Health data regulated by HIPAA can be repurposed: prescription data originally gathered to ensure access to patient records in emergency care settings is being employed by pharmacies that have begun to incentivize patients to waive their HIPAA rights and sell their data to insurance companies. Similarly, insurers and employers may incentivize participation in questionnaires or wellness programs that enable the collection of health information.

Health data generated outside of “covered entities” are not protected by HIPAA, but could have limited protection from other laws. The Americans with Disabilities Act of 1990 (ADA), for example, “prohibits employers from administering medical examinations and making other disability-related inquiries” unless the inquiry is job-related. Similarly, the Genetic Information Nondiscrimination Act of 2008 (GINA) prohibits employers from discriminating on the basis of genetic information, and makes it illegal to “request, require, or purchase genetic
Both laws prohibit a concerning type of discrimination but are limited in scope to certain types of information in certain contexts.

Publicly funded medical research is also subject to the Federal Policy for the Protection of Human Subjects, also known as the Common Rule. The Belmont Report, on which the Common Rule is largely based, was developed to provide ethical standards for research in response to the history of unethical experiments in the US, including the US Public Health Service Syphilis Study at Tuskegee. The Common Rule creates requirements for informed consent, special protections for members of vulnerable groups, and oversight by Institutional Review Boards—empaneled stakeholders tasked with ensuring that proposed research protects human subjects according to “ethical principles and federal regulations.” The Common Rule also mandates that medical data collection be transparent to patients, secure, and restricted in use and scope, in stark contrast to data practices in the corporate sector that are more opaque to consumers.

In addition to legislation, federal regulatory bodies play a role in overseeing and ensuring health data privacy. The US Food and Drug Administration (FDA) regulates medical devices, but has resisted regulating most mobile medical applications because the agency does not regard these apps as posing a risk to a patient’s safety when functioning as intended. (Though some have argued the potential of these apps to provide inaccurate medical data should require FDA oversight.) Additionally, the Federal Trade Commission (FTC), which oversees antitrust enforcement and consumer protection, and the National Telecommunications and Information Administration (NTIA), which advises the President on telecom issues, have largely resisted regulating data collection and have instead released guidelines by which companies are expected to self-regulate if unintended health data is a byproduct of their work. Existing US laws were not designed to protect the broad stores of health-relevant data produced today originating from outside of “covered entities,” and so they may not be the most effective at protecting it. Simultaneously, existing institutions have not stepped in to regulate health-relevant data.

**Beyond the Bounds of Healthcare Regulation**

The limits of current regulation and legislation are striking when the breadth of interest in health-relevant data, both within and beyond the medical and public health sectors, is made explicit. Some technologies produce data about health, including mobile phone apps, wearable devices, and some “Internet of Things” (IoT) devices. These devices may record self check-ins, self-assessments, location, and sleep patterns, which may capture sensitive information about both physical health and mental health. Sensors in wearables and some IoT devices may capture additional physiological data, such as Sleep Number beds that measure heart rate, breathing, and movement. Other data may also be revealing about health: information from
home testing labs or shared with gyms; correspondence with disease advocacy groups; and information disclosed to nutritional counselors, alternative medicine practitioners, cosmetic services providers, massage therapists, and marketers of non-prescription health products all fall outside the scope of HIPAA. Multipurpose systems may also capture health-relevant information that can be gleaned from social network data and search data. For example, researchers have already published a dataset of more than 123 million tweets about Covid-19 to study conversation dynamics and misinformation, and used search trends to study the effect of local cases of the pandemic on information-seeking behavior. Information that is not explicitly about health may nevertheless support inferences about health. Data about a person’s physical activity coupled with data about income, race, ethnicity, and neighborhood, for example, may predict risk of cardiovascular disease.

Plans from the late 1990s for at-home patient monitoring that restricted data access to healthcare institutions underscore that the broad dissemination of modern health-relevant data collection was not inevitable. The healthcare sector has adopted new technologically enabled healthcare practices since the passage of HIPAA, including telehealth, that are still covered by the law. (However, as part of the Coronavirus Preparedness and Response Supplemental Appropriations Act passed on March 6, 2020, HIPAA privacy rules were temporarily waived to allow doctors and patients to connect via video conferencing apps like FaceTime and Zoom.)

Actors interested in health data represent a diverse set of institutions and perspectives, including: healthcare systems, health insurers, disease-specific advocacy groups, academics, and professional societies. Through the allure of mission creep, public health institutions may collect and repurpose data under HIPAA protection so extensively they begin to take the form of “general searches, which the authors of the Bill of Rights were so concerned to protect against” to borrow the words of caution voiced by the US Department of Defense’s Technology and Privacy Committee following September 11, 2001. HIPAA may be too narrow not only for modern technologies, but for modern data economies, where the share of health data covered by HIPAA is small given the information that can be collected and traded beyond its originally intended scope.

Health Data and Privacy in the European Union and United Kingdom

Many European countries regulate data use practices generally, rather than establishing sector-by-sector standards. Data protection law began to develop in Europe in the 1970s, and was formalized by the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (Convention 108) in 1981 by the Council of Europe, and the Data Protection Directive (Directive 95/46/EC) by the European Parliament and Council in 1995. The Data Protection Directive was replaced by the General Data Protection Regulation (GDPR), which was passed in 2016 and took effect on May 25, 2018.
GDPR applies to personal data processed in a wide variety of public and private entities and aims to make data use clear and based on consumer opt-in and consent. GDPR adopts a risk-based, context-specific approach and, apart from certain exemptions for research, requires case-by-case assessment for data use beyond its initial purpose. The United Kingdom concurrently developed its own data regulation in the Data Protection Act (DPA). The DPA was originally passed in 1984 and replaced in 1998 to legally implement Directive 95/46/EC, and was amended in 2018 to legislate data protection in the UK given the UK’s transition out of the European Union. A number of additional regulations address privacy and confidentiality for medical institutions across the UK. EU member countries may similarly enact their own legislation to clarify the applicability of GDPR or further define the scope of its exceptions, including for genetic, biometric, or other health-related data. GDPR is still in effect in the UK through the transition from the EU until the end of 2020, and the UK government plans to continue the regulation after that period.

Despite similar pressure to commercialize health data in Europe, use of health data has taken a different form under GDPR and DPA. GDPR provides exceptions for sharing health data without explicit patient consent for medical treatment, scientific research, and public health, but otherwise the processing of “data concerning health” is prohibited without explicit consent. Notably, this regulation largely prevents the accumulation and commercial use of non-healthcare health data, as it not only applies to the processing of personal data generally, but also to physical or physiological data that may not directly identify a person or implicate health status, but that could make a person identifiable. Before GDPR went into effect, the DPA prevented such data misuse in England. In 2015, Google DeepMind entered into a partnership with a National Health Service (NHS)-based entity to obtain 1.6 million identifiable medical records to test an app that would help monitor kidney injury and failure. The data was transferred under an “implied consent” rule that allows the NHS to share patient information without consent for direct patient care, but was later condemned by the National Data Guardian for the English Department of Health and the Information Commissioner for the UK for violating data protection rules given that the data was not used for direct patient clinical benefit, and the information shared was not justified under DPA. There is ongoing interest in applying data science and AI to healthcare in the UK, but laws governing how any information related to health can be acquired and used are more robust to context in the UK and EU than in the US. Moreover, the UK and EU context-agnostic approach better mitigates regulatory lag.

Is Covid-19 a Pivot-Point, or is it Amplifying Existing Processes?
Digital health surveillance evolved in tandem with the internet. For example, ProMED-mail was a precursor to early outbreak notifications that used emails about emerging infectious diseases in the early 1990s. Digital health surveillance also became possible
through technological transformations within healthcare institutions, like the introduction of electronic health records intended to improve the breadth, timeliness, and completeness of case reporting. Some public health surveillance systems, such as systems developed to detect foodborne illness via Yelp reviews by the City of New York and via tweets by the City of Chicago, adopt internet-based data like search, review, and social media posts, though additional evaluation is needed to determine if they add value to established methods. Health experts are especially interested in using internet-based data for “nowcasting” methods to predict real-time or immediately expected disease incidence or prevalence, or using computing techniques for long-term predictive analytics that may enable physicians to screen and recommend preventive measures earlier. Like other data science systems, these public health surveillance systems are enabled by increased data availability, storage capacity, and computing power. These systems focus on a variety of impacts to improve public health, including identifying early outbreaks, stewarding rapid action, measuring the burden of a disease, guiding the implementation and evaluation of programs and policy, and studying the effects of changes in health practices.

Commercial entities have expanded their capacities to collect health data and make inferences from it. Some data, like self-assessments and biometric data, support very direct inferences about health. Companies including LexisNexis, Facebook, and other advertising firms have also developed classifiers to make additional inferences from their data, including detecting emotional and mental health indicators, and to predicting “life outcomes,” including substance use, depression, and suicide. During the Covid-19 pandemic, researchers have developed apps to predict personal and regional risk based on user-reported information and networks, while others have used Facebook's Data for Good program to measure the success or failure of lockdown efforts using aggregate GPS location data.

The collection and use of health data outside of health institutions has a variety of consequences. For example, advertising has long tugged at audiences' heartstrings to sell products, but using emotional analytics in digital advertising could mean personalized emotional manipulation that preys on health traits. Governments may also be interested in capturing biometric data and using it creatively: in the US, Fitbits and pacemakers have been used by police in legal cases. During the Covid-19 pandemic, the Chinese government has used information including health status to assign codes that indicate individual infection risk. Officials are considering expanding the app to include a “personal health index” rank based on how much users sleep and how many steps they take, among other data points.

As varied forms of information that might help mitigate Covid-19 are marshaled, both within and without health data privacy regulations, it is important to highlight the limitations of this data. Many concerns of adopting data science approaches to health monitoring apply
to institutions both in traditional public health and corporations. Adequately de-identifying data is increasingly challenging, and computer systems may be compromised by mundane challenges like software bugs, computer shutdowns, user error, and security vulnerabilities. Non-traditional health data may also be biased in a number of ways, including (1) undersampling or excluding members of communities who could benefit the most, and (2) capturing public awareness of phenomena rather than disease occurrence. These systems also tend to be non-transparent, hindering efforts to hold powerful data brokers accountable and prepare for the implications of their data use. Given that there is some evidence people may self-censor when their personal health information is stored digitally, the accuracy of the data may be unreliable. It is also difficult to ensure autonomy to the people reflected by data as it is prohibitively difficult to obtain informed consent for subjects in very large datasets. It can be impossible for subjects to anticipate how their data will be used and, because analysis necessarily enables predictions about people beyond the sample, broader questions arise about privacy both for individuals but also for communities and groups. Overreach, commercialization of data, and centralization of power are also of concern, as data, computing resources, control over technology, and talent may encourage collaboration between companies, academics, and publicly funded health institutions—but doing so creates ambiguity in mission and the potential for malfeasance. “Mission creep” in response to rich data stores is tempting for state and non-state institutions alike.

What Does Health Privacy Teach Us about Surveillance?
The pattern of initiatives for health and science being repurposed is persistent. Early sensors were developed partly to monitor animal populations. When sensors were repurposed for human wearables beginning in the 1990s, a prominent researcher hoped they would serve a positive purpose by improving one’s senses, memory, social life, and composure. Today, wearables analyze and monetize intimate physiological signals. During the first decade of the 21st century, researchers at MIT developed technology to assist autistic children in recognizing human emotion, and that technology was eventually repurposed for emotion recognition in advertising. Repurposing scientific development for financial gain, though, is not inevitable: human genetics and predictive data analytics are similar in their predictive accuracy, privacy and discrimination issues, and potential for psychological harm. Yet applied ethics research was funded alongside genetics research, and policies to protect genetic information were eventually enacted through the passage of GINA in 2008 and an amendment to HIPAA in 2013. These historical precedents raise questions about which path today’s Covid-19-related innovations will follow and whether predictions that today sound fanciful (e.g., will corporate actors attempt to replace, rather than assist, doctors with predictive algorithms?) will become tomorrow’s reality.
To what extent is health data “special data?” In 1994 and again in 2001, the Institute of Medicine argued that the usual mechanisms of privacy governance (corporate ethics, consumer pressure, and “occasional attention by the government”) were not sufficient for health data. Today, health data exceptionalism is encoded in US law: the country has sector-based privacy laws, and the regulation of health data through HIPAA provides far and away the strongest data protection of any sector. Yet the traditional arguments about what make medical data exceptional—the breadth of personal information one shares in psychotherapy, or the predictive capacity of genetic data—are also features of non-health data in today’s data economies. The characteristics that make it difficult to differentiate the exceptional characteristics of health data are also characteristics that make it difficult to differentiate between health and non-health data. Not only because sophisticated technology can collect fine-grained physiological data and predict health from non-medical data, but also because companies develop unsophisticated methods to collect non-protected health information. The right to health is concerned not only with the availability of healthcare services, but also with underlying determinants of health. If privacy and confidentiality are important for the realization of good health, it is imperative to begin to account for the ways our health is reflected in many features of our lives, not just in our clinical records.

Clara Hanson
Alondra Nelson

August 2020
ENDNOTES


4 Backman et al., “Health Systems.”


9 Backman et al., “Health Systems.”


18 Dixon and Gellman, *Scoring of America*; Francis and Francis, “Informatics and Public Health.”


21 Pasquale, *The Black Box Society*.


24 Haskins, “Wearable Technology.”


26 Cohen and Mello, “HIPAA.”


29 Cohen and Mello, “HIPAA.”

30 Emamian et al., “Regulating Mobile.”

31 Zuboff, *Age of Surveillance Capitalism*.

32 Dixon and Gellman, *Scoring of America*.


36 Cohen and Mello, “HIPAA.”

37 Zuboff, *Age of Surveillance Capitalism*.


39 Seabrook, “Promise and the Peril.”

40 Francis and Francis, “Informatics and Public Health.”

42 Cohen and Mello, “HIPAA.”


46 Carey, *Data Protection*.

47 For a general overview, see *Confidentiality: Key Legislation* (General Medical Council, n.d.).

48 Carey, *Data Protection*.


51 Price et al., “Shadow Health Records.”


Aiello et al., “Social Media.”


Aiello et al., “Social Media.”


Groseclose and Buckeridge, “Public Health Surveillance.”

Groseclose and Buckeridge, “Public Health Surveillance.”


Zuboff, *Age of Surveillance Capitalism*.

Zuboff, *Age of Surveillance Capitalism*.


72 Seabrook, “Promise and the Peril.”

73 Aiello et al., “Social Media.”


79 Zuboff, *Age of Surveillance Capitalism*.

80 Zuboff, *Age of Surveillance Capitalism*.

81 Zuboff, *Age of Surveillance Capitalism*.


83 Pasquale *The Black Box Society*.

84 Seabrook, “Promise and the Peril.”
85 Terry, “Big Data Proxies.”

86 Terry, “Big Data Proxies.”