

Placing Equity

at the Center
of Health Care
& Technology



The **Center for Democracy & Technology** (CDT) is a 25-year-old 501(c)3 nonpartisan nonprofit organization working to promote democratic values by shaping technology policy and architecture. The organisation is headquartered in Washington, D.C. and has a Europe Office in Brussels, Belgium.

Placing Equity at the Center of Health Care & Technology

Author

Andrew Crawford

Acknowledgements

This report and related framework are both made possible with the support of the Robert Wood Johnson Foundation, and with assistance from Executives for Health Innovation (EHI), our steering committee, working groups, and new phase-two participants. Many thanks to all for their invaluable engagement, help, and guidance.

Footnotes in this report include original links as well as links archived and shortened by the [Perma.cc](#) service. The Perma.cc links also contain information on the date of retrieval and archive.


March 2022

Table of Contents

Introduction	5
Current Legal Regime for Health Data in the United States Leaves Big Gaps	7
The Gaps in Regulation Create Avenues for Harmful Data Practices	11
Exposure, Embarrassment, and Stigma	11
Inaccurate and Biased Data	12
Lack of Autonomy	13
Discriminatory Health Treatment	14
Lack of Trust in Technology and Health Services	17
The COVID-19 Pandemic Highlighted and Heightened Disparities While Also Increasing the Demand for Technology and Data	19
Ensuring Equitable Data Practices for Health Data	23
New Definition for Health Data	23
Social Determinants of Health Data is Consumer Health Information	25
Appropriate Collection and Use of Data About Race or Sexual Orientation May Help Address Long-Standing Inequities	26
Appropriate Collection and Use of Data that Reflects Disability May Help Address Long-Standing Inequities	27
Better Research Practices Using Consumer Health Information	28
Diversity of Data	30
Conclusion	32

Introduction

Data that may not initially appear to have any relation to our health can be analyzed, combined with other data, and used in ways that reveal a lot about our physical and mental health.



Modern consumer-facing technology runs on data. Every moment, we as consumers produce data that is collected and used by technology around us. Modern consumer health technologies are able to collect and store troves of individualized health information. In fact, data that may not initially appear to have any relation to our health can be analyzed, combined with other data, and used in ways that reveal a lot about our physical and mental health.

When deployed, shared and used appropriately, data generated by new and existing technologies has the potential to help us all be healthier and address a history of inequities, including those in the provision of physical and behavioral health care. However, for these critical, indeed life-changing, benefits to be realized, we must change certain practices surrounding how consumer health data is collected, shared, and used. In particular, additional privacy safeguards are necessary to ensure that everyone benefits from, and enjoys the same protections and opportunities, including consumers and people from marginalized communities.

People continue to want data about their health protected and recognize the stakes when it is not. A Recent Pew Charitable Trusts survey found that patients want their health data protected and kept private. When “[a]sked about the privacy of their health data when it is downloaded to apps, 35% of respondents said they were extremely or very concerned; that number rose to 62% when they were told that federal privacy laws, such as HIPAA, do not cover data downloaded to apps, and that the apps’ terms of service would protect the data instead.”¹ Additionally, the

1 Pew Charitable Trusts, Most Americans Want to Share and Access More Digital Health Data (2021), https://www.pewtrusts.org/-/media/assets/2021/07/americans_support_federal_efforts_v5.pdf [<https://perma.cc/Q24N-KY69>].

Pew study details how respondents “...expressed particular worries about identity theft and blackmail; discrimination; the absence of federal data protections; and a desire to protect their information from access by large technology companies.”²

Reforms around health data must be rooted in fair and equitable principles. This paper aims to identify and address ways in which privacy protections that apply to consumer health information can truly benefit everyone, including communities that have been, and continue to be, underrepresented, overlooked, and harmed by current health data practices. This paper begins by briefly outlining gaps in existing privacy protections and identifying classes of consumer health information that are under-protected here in the U.S., or are protected differently depending on which entity is processing them. Next, it discusses long-standing inequities in the use of data and technology for health care delivery. Finally, it sets forth potential reforms in the collection and processing of data that can help ensure greater equity in health care.

2 *Id.*

Current Legal Regime for Health Data in the United States Leaves Big Gaps

Health data is not regulated by a single law or standard in the U.S. The primary privacy law that protects certain health data is the Health Insurance Portability and Accountability Act of 1996 (HIPAA).

HIPAA and its accompanying regulations establish rules regarding how health data should be collected, retained, shared, and used. Importantly however, HIPAA's reach is limited and focuses its protections on specific holders of data, referred to as "covered entities." HIPAA's data protections cover any information created or received by a covered entity that relates to the past, present, or future physical or mental health or condition of an individual, the provision of health care to an individual, or payment for the provision of health care to an individual.³ It also covers such data when shared with "business associates" of covered entities.

While the definitions and protections encapsulated in HIPAA provide some protections for health data in the hands of covered entities,⁴ the world did not anticipate the technological realities of today when the law was passed in 1996. Indeed, back in the 1990s when HIPAA was drafted,

3 Health Insurance Portability and Accountability Act of 1996, P.L. 104-191, 110 Stat. 1936, August 21, 1996.

4 HIPAA and its accompanying privacy rule "limits covered entities from using protected health information or sharing it with third parties without valid patient authorization, unless the use is for purposes of treatment, payment, or "health care operations," or falls within a specific statutory exception." Congressional Research Service, COVID-19: Digital Contact Tracing and Privacy Law (2020), <https://crsreports.congress.gov/product/pdf/LSB/LSB10511>.

debated, and passed, legislators were focused on digitizing and standardizing electronic payments and the internet had yet to venture outside our desktop computers and into our pockets – the first iPhone was not available until 2007.

In the time since HIPAA was established, technology capable of collecting, processing, retaining, and sharing data about us has exploded in volume and capabilities. A host of entities outside of the traditional health care space generate, collect, share, process, and retain health data. Nearly 100,000 medical, health, and fitness smartphone apps are available in the Google Play and Apple App stores.⁵ In addition, connected “Internet of Things” (IoT) devices, wearables, fitness trackers, and websites all generate, share, and use health data. Such data often is not covered by HIPAA’s rules because those holding the data are outside of HIPAA’s limited list of covered entities and are not acting as business associates.

The complexity does not end there. Some products can have certain features governed by HIPAA while other functions fall outside of it. Take, for example, services that allow people to make medical appointments online.⁶ Data collected by some of these services to make an appointment with a doctor may be protected by HIPAA.⁷ However, the same protections may not extend to other user actions within the same services, like a person’s searches for doctors, surveys, or medical history forms that are not associated with a particular medical provider.⁸

Moreover, the line between covered and non-covered entities is permeable. For example, recent actions by the Office of the National Coordinator for Health Information Technology within HHS have given patients greater access to their medical records held by HIPAA-covered entities.⁹ As a result, patients will be able to easily access and control data about their health and store that information in a place of their choosing.¹⁰ Many of those places, such as convenient smartphone apps, are not HIPAA-covered entities and thus moving data into these places would mean moving it outside the coverage of HIPAA. Conversely, data that originates with apps or other entities not subject to HIPAA that is then transferred to a covered entity for health purposes may become subject to HIPAA.

5 Robby Berman, *Do mHealth apps protect user privacy?*, Medical News Today (June 21, 2021), <https://www.medicalnewstoday.com/articles/do-mhealth-apps-protect-user-privacy?c=1381705972936> [<https://perma.cc/R6V9-YKLQ>].

6 Adam Tanner, *Find a Doctor Near You? Yes, but Medical Booking Sites Have Downsides, Too.*, Consumer Reports (June 25, 2021), <https://www.consumerreports.org/consumer-protection/medical-booking-sites-have-downsides-a7415010250/> [<https://perma.cc/ZSJ7-6MST>].

7 *Id.*

8 *Id.*

9 The Office of the National Coordinator for Health Information Technology, *ONC’s Cures Act Final Rule supports seamless and secure access, exchange, and use of electronic health information.*, <https://www.healthit.gov/curesrule/> [<https://perma.cc/CSE4-FJ26>].

10 *Id.*

Data that at one moment is controlled by a HIPAA-covered entity can quickly move to a non-HIPAA-covered entity and vice versa. Data created and collected outside the traditional health system routinely finds its way to HIPAA-covered entities.

These technical distinctions and complexities, which may be explained in privacy policies and terms that many people do not read, can allow a company to share or even sell would-be patients' data, including search query information like doctor names or medical conditions that are revealing about peoples' health.¹¹ When selecting and using IoT products and apps, most people don't regularly contemplate when HIPAA privacy protections apply and when they do not.

Health information that falls outside of HIPAA generally will still be within the authority of the Federal Trade Commission (FTC), which regulates unfair or deceptive acts or practices. Specifically, under Section 5 of the FTC Act, the agency can bring enforcement actions against an entity, "for collecting or using personal information in a deceptive or unfair manner, such as when a company's privacy practices contradict its posted privacy policy."¹² The FTC has on occasion used this authority to bring actions against consumer health tech products whose data practices harm consumers.¹³ But the FTC's Section 5 authority is limited and does not provide comprehensive privacy protection.

The health data ecosystem is quickly transforming. Data that at one moment is controlled by a HIPAA-covered entity can quickly move to a non-HIPAA-covered entity and vice versa. Data created and collected outside the traditional health system routinely finds its way to HIPAA-covered entities. New and emerging tech in both the consumer and traditional health care space will continue to increase the frequency, amount, and quality of health data that is sharable. For instance, tech like wearable health devices have become indispensable tools for aiding in medical research and encouraging people to develop healthier habits.¹⁴ We

¹¹ Tanner (2021), *supra* n. 6.

¹² Congressional Research Service (2020), *supra* n. 4.

¹³ For example, in early 2021, the FTC brought an action against a mobile health app designed to allow users to track their reproductive health. Flo Health Inc., FTC Docket No. C-4747 (2021), <https://www.ftc.gov/legal-library/browse/cases-proceedings/192-3133-flo-health-inc> [<https://perma.cc/PX4H-64A2>].

¹⁴ Lisa Eadicicco, *Fitbit and Apple know their smartwatches aren't medical devices. But do you?*, CNET (Jan. 14 2022), <https://www.cnet.com/tech/mobile/features/fitbit-apple-know-smartwatches-arent-medical-devices-but-do-you/> [<https://perma.cc/4TBQ-WSQJ>].

are witnessing more crossover between medical and consumer devices and the health data they generate.¹⁵ As commercial wearables gain more advanced health monitoring capabilities, medical device makers are attempting to reach broader audiences, and we are witnessing those two streams coming together.¹⁶ There is more and more overlap between the traditional consumer-facing product universe and health providers.¹⁷

This overlap can empower patients and drive better health outcomes. But inequitable data practices found in both the consumer and traditional health care spaces have not gone away. As health data continues to be collected, shared, and used, we must confront inequitable practices and prevent them from further entrenching themselves and harming people, especially from communities that have been, and continue to be, marginalized.

15 Lisa Eadicicco, *Fitness trackers are getting more personal, powerful in 2022 and beyond*, CNET (Jan. 29, 2022), <https://www.cnet.com/tech/mobile/fitness-trackers-are-getting-more-personal-powerful-in-2022-and-beyond/> [<https://perma.cc/TZM4-LTNS>].

16 *Id.*

17 *Id.*

The Gaps in Regulation Create Avenues for Harmful Data Practices

Inadequate data protections result in a greater likelihood that consumers will be harmed when their health data is misused in ways they don't anticipate, expect, or even know about. Look no further than a recent International Digital Accountability Council (IDAC) report, which found that, while most consumer-facing health apps they examined “are observing the letter of the laws and platform terms that they are required to follow, ... some widely-used apps fail to meet even basic platform requirements because they send unencrypted user data, have inadequate or missing privacy policies, or collect granular information about user location without adequate explanation.”¹⁸ The IDAC report also found that “[t]he majority of apps investigated have questionable practices and disclosures around third-party data sharing, illustrating a clear mismatch between current legal protections and the widespread collection and sharing of sensitive health information.”¹⁹

Unregulated or inappropriate data use, like those documented in the IDAC report, can produce biased data, compound historical discrimination, and yield incorrect assumptions. Unfortunately, all too often, these risks are disproportionately borne by historically marginalized groups, including people of color, immigrants, Indigenous populations, women, people with disabilities, and the LGBTQ+ community. The resulting harms can take a number of different forms.

Exposure, Embarrassment, and Stigma

Health-related data collected, shared, and used by consumer-facing tech can be extremely personal and sensitive, and embarrassing if released. Examples may include data about conditions that are especially sensitive because of accompanying, unwarranted social stigmas and discrimination. These harms frequently are felt by people from marginalized communities.

18 International Digital Accountability Council, *Digital Health is Public Health: Consumers' Privacy & Security in the Mobile Health App Ecosystem* (2021), <https://digitalwatchdog.org/wp-content/uploads/2022/01/Digital-Health-is-Public-Health-Consumers-Privacy-and-Security-in-the-Mobile-Health-App-Ecosystem.pdf>. [<https://perma.cc/4FZX-JH6V>].

19 *Id.*

For example, disclosure of information that reveals “an HIV diagnosis or LGBTQ+ identity can be highly stigmatizing and often lead to discrimination, or even violence.”²⁰ In 2018, it was revealed that Grindr, a dating app geared towards the LGBTQ+ community, “had provided users’ HIV status and GPS location data, along with other profile details including email addresses, to two companies hired to test the app’s technical performance.”²¹ This type of sensitive data sharing by Grindr was inconsistent with user expectations and could result in real harm.²² Grindr’s data practices meant that a user who wished their information to stay within Grindr instead had their information shared with outside parties. A user could be outed and subsequently stigmatized even though they believed their data would be used for the specific and limited purposes within the app.

Inappropriately sharing and exposing reproductive health information can also be embarrassing and harm users. A recent FTC case alleged that Flo Health “shared the health information of users with outside data analytics providers after promising that such information would be kept private.”²³ In its complaint, the FTC alleged that, by “encouraging millions of women to input extensive information about their bodies and mental and physical health, [Flo] has collected personal information about consumers, including name, email address, date of birth, place of residence, dates of menstrual cycles, when pregnancies started and ended, menstrual and pregnancy-related symptoms, weight, and temperature.”²⁴ Because it misrepresented how it handled users’ reproductive health data, Flo was required to notify its affected users and instruct any third party that received users’ health information to destroy that data.²⁵

Inaccurate and Biased Data

As discussed above, data can flow between consumer-facing technology and the traditional health system. If data from consumer-facing tech is being used for health purposes like diagnosis or access to benefits, the data needs to be correct. Inaccurate data can disproportionately harm those whose data is not accurately or fully represented and result in

20 Alison Bateman-House, *Why Grindr’s Privacy Breach Matters To Everyone*, Forbes (Apr. 10, 2018), <https://www.forbes.com/sites/alisonbatemanhouse/2018/04/10/why-grindr-privacy-breach-matters-to-everyone/?sh=9ac72a767f40> [<https://perma.cc/VT55-YNC8>].

21 *Id.*

22 *Id.*

23 FTC (2021), *supra* n. 13.

24 Flo Health Inc., FTC Docket No. C-4747 (2021), https://www.ftc.gov/system/files/documents/cases/192_3133_flo_health_complaint.pdf [<https://perma.cc/DVG9-SGJA>].

25 Flo also is prohibited from misrepresenting: (1) the purposes for which it (or entities to whom it discloses data) collect, maintain, use, or disclose the data; (2) how much consumers can control these data uses; (3) its compliance with any privacy, security, or compliance program; and (4) how it collects, maintains, uses, discloses, deletes, or protects users’ personal information.” Flo Health Inc., FTC Docket No. C-4747 (2021), <https://www.ftc.gov/news-events/news/press-releases/2021/06/ftc-finalizes-order-flo-health-fertility-tracking-app-shared-sensitive-health-data-facebook-google> [<https://perma.cc/MG6D-RNB4>].

negative health outcomes and lost or denied services and benefits. Moreover, regardless of the source of data, certain data practices continue to rely on inaccurate, unrepresentative, or biased information that harm people from underrepresented and overlooked communities.²⁶

Unrepresentative data sets, sometimes originating from sources such as consumer wearables, are frequently used and relied upon by predictive health technologies. For instance, Wired reported that “skewed data sets are the norm in health AI research, due to historical and ongoing health inequalities.”²⁷ That story cited a 2020 study by Stanford researchers that found 71 percent of data used in studies that applied deep learning to U.S. medical data came from California, Massachusetts, or New York, with little or no representation from the other 47 states.²⁸ The same story also noted how a recent review of more than 150 studies using machine learning to predict diagnoses or courses of disease concluded that most “show poor methodological quality and are at high risk of bias.”²⁹ Biased, inaccurate, and incomplete data, regardless of whether it comes from the consumer or health care space, can be harmful to people from underrepresented and overlooked communities.

Lack of Autonomy

When health data is shared with others, especially for unanticipated uses, people can quickly lose control over it. Acute harms can result when data purportedly used for one purpose is used in other ways that harm communities.

Consumer health apps, including menstrual cycle tracking apps, often collect, share, and sell consumer health data to third parties like advertisers, insurers, and tech companies.³⁰ For example, recent news articles exposed how the nonprofit mental-health hotline Crisis Text Line ended a data-sharing relationship with a for-profit spinoff only days after reports detailing its troubling data practices emerged.³¹ People use this service to seek help for problems such as suicidal thoughts, anxiety and emotional abuse and in so doing disclosed

26 See generally, American Heart Association, Considerations for Cardiovascular Genetic and Genomic Research With Marginalized Racial and Ethnic Groups and Indigenous Peoples: A Scientific Statement From the American Heart Association (2021), <https://www.ahajournals.org/doi/10.1161/HCG.0000000000000084> [<https://perma.cc/7UGK-EETY>].

27 Tom Simonite, *When It Comes to Health Care, AI Has a Long Way to Go*, Wired (Jan. 16, 2022), <https://www.wired.com/story/health-care-ai-long-way-to-go/> [<https://perma.cc/79PV-SFW9>].

28 *Id.*

29 *Id.*

30 Emily Kwong, *When Tracking Your Period Lets Companies Track You*, NPR (Jan. 18, 2022), <https://www.npr.org/transcripts/1068930998> [<https://perma.cc/SW8V-EUTV>].

31 John Hendel, *Crisis Text Line ends data-sharing relationship with for-profit*, Politico (Jan. 31, 2022), <https://subscriber.politicopro.com/article/2022/01/crisis-text-line-ends-data-sharing-relationship-with-for-profit-spinoff-00004001?source=email>

highly personal and sensitive information.³² Yet, no law or regulation prohibited the sharing of this information for secondary purposes unrelated to the help that the individuals were seeking. The swift outcry and change of course demonstrate how people do not want their sensitive mental health data used for unrelated product development.

Consumer-facing mental health apps are not the only ones sharing consumer health information. Data about expecting parents can be very valuable as they change their spending habits to prepare for their new arrival.³³ While many people may find the subsequent resulting tailored advertisements useful, others may not.³⁴ Certain data practices limit or remove individual autonomy and can also creep into other areas of life. For example, some consumer health data is shared with others, like an insurer or employer, and can have real-life impacts on access to a job and/or health care.³⁵

Autonomy can also be lost when an app, wearable, or any other consumer-facing piece of technology is not accessible to individual users - forcing them to disclose sensitive health information to others or simply not use certain services or technologies. For example, people with disabilities may not be able to utilize certain services and products independently if they are not accessible. Instead, to use a poorly designed product, they may be required to reveal their personal health information to another person who can then relay information or enter and submit the information on their behalf. When the needs of certain communities go unrecognized or ignored, products and services' lack of access result in some consumers' inability to independently access or transmit their health data without another person helping, with a resulting loss of autonomy and privacy.

Discriminatory Health Treatment

When not appropriately used, data can exacerbate existing biases in the provision of care and treatment. Data use that relies on biased and non-representative foundations will continue to disproportionately harm underrepresented and overlooked communities.

Long-standing bias, racism, and inequities have harmed minority and marginalized persons and communities for centuries leading to alarming disparities in health outcomes and healthcare access in communities of color, low-income communities, the LGBTQ+ community, the disability community, and others. And because data flows between

³² *Id.*

³³ "A pregnant person's health data is worth 15 times more than a nonpregnant person's because companies know that soon-to-be parents are opening their wallets." Kwong (2022), *supra* n. 30.

³⁴ Kashmir Hill, *How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did*, *Forbes* (Feb. 16, 2012), <https://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/?sh=4f0f556b6668> [<https://perma.cc/X7H2-BM6A>].

³⁵ Drew Harwell, *Is your pregnancy app sharing your intimate data with your boss?*, *Wash. Post* (Apr. 10, 2019), <https://www.washingtonpost.com/technology/2019/04/10/tracking-your-pregnancy-an-app-may-be-more-public-than-you-think/> [<https://perma.cc/Z5VK-BE99>].

Researchers have found that a widely used algorithm deployed to guide care decision-making for millions of people by predicting which patients will benefit from extra care dramatically underestimated the health needs of the sickest Black patients.

consumer-facing tech and tech used in the traditional health care space, inequitable data practices found in one can easily move with data into the other.

Examples of disparities are not hard to find. A recent *Slate* piece explores overtly racist practices incorporated into medical algorithms frequently used in the treatment and care of patients with kidney disease. The article notes, "...racialized tropes have been subtly knit into medical algorithms and are buried in the decision-making of doctors every day."³⁶ Specifically the article details how the mathematical equation and data used by medical algorithms to determine how sick kidney patients are give each patient a score, called eGFR score.³⁷ This score determines eligibility not just for a patient's placement on the transplant list, but also the stage of care at which they can access Medicare-covered nutritional therapy and reimbursed kidney disease treatment, and referral to a doctor that specializes in kidney disease.³⁸

Black patients are harmed by this process because their data and subsequent eGFR scores are calculated using a different and arbitrary metric compared to other non-Black patients.³⁹ This multiplier inflates Black patients' scores, artificially creating the appearance that Black patients with kidney disease are healthier than they truly are.⁴⁰ As a result, Black patients have to reach more advanced stages of kidney disease before they are considered sick enough to qualify for certain treatments or interventions, which delays or limits the critical care they receive.⁴¹

Unfortunately, the inequitable treatment of kidney patients is not the only example of bias and harmful tech-assisted

36 Jennifer Tsai, *Jordan Crowley Would Be in Line for a Kidney - if He Were Deemed White Enough*, *Slate* (June 27, 2021), <https://slate.com/technology/2021/06/kidney-transplant-dialysis-race-adjustment.html> [<https://perma.cc/3BK6-HQK9>].

37 *Id.*

38 *Id.*

39 *Id.*

40 *Id.*

41 *Id.*

approaches to treatment. Researchers have found that a widely used algorithm deployed to guide care decision-making for millions of people by predicting which patients will benefit from extra care dramatically underestimated the health needs of the sickest Black patients.⁴²

The algorithm “...used a seemingly race-blind metric: how much patients would cost the health-care system in the future.”⁴³ However, by focusing on costs, the algorithm produced biased results.⁴⁴ The algorithm failed to account for the fact that Black patients “...incurred about \$1,800 less in medical costs per year than white patients with the same number of chronic conditions; thus the algorithm scored white patients as equally at risk of future health problems as black patients who had many more diseases.”⁴⁵

In essence, while assigning higher risk scores in algorithms to patients who have higher health care costs may have seemed reasonable to the developers because higher health costs are often associated with greater needs, this methodology failed to account for the systemic and long-standing inequities in care that have resulted in fewer expenditures on Black patients. This resulted in further inequitable care, excluding Black patients from care management programs that dedicate additional resources to coordinate care for higher risk programs.

Once the researchers detected this bias, they worked with the major health services company that produced this particular algorithm to correct the problem.⁴⁶ The researchers who looked into this specific case believe “the same issue almost certainly exists in other tools used by other private companies, nonprofit health systems and government agencies to manage the health care of about 200 million people in the United States each year.”⁴⁷

These harms can compound and metastasize. For example, if certain data collection mechanisms are inaccessible as discussed above, secondary harms can also result when data from and about certain communities fails to be collected, resulting in products and services that rely on incomplete data and may not meet the needs of all users. Apps that are inaccessible to certain populations and individuals due to their design or platform structure run the risk of isolating communities, keeping data about them from being used in appropriate and beneficial ways. Health tech products may also fail to test, involve, and consider certain groups when they design and offer certain products because developers

42 Carolyn Y. Johnson, *Racial Bias in a Medical Algorithm Favors White Patients Over Sicker Black Patients*, Wash. Post (Oct. 24, 2019), <https://www.washingtonpost.com/health/2019/10/24/racial-bias-medical-algorithm-favors-white-patients-over-sicker-black-patients/> [<https://perma.cc/6CY4-9RB9>].

43 *Id.*

44 *Id.*, See also, Science, *Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations* (2019), https://www.ftc.gov/system/files/documents/public_events/1548288/privacycon-2020-ziad-obermeyer.pdf [<https://perma.cc/VF8C-4B35>].

45 Johnson (2019), *supra* no. 42.

46 *Id.*

47 *Id.*

fail to adequately consider the needs of underrepresented communities who will use their products.

Lack of Trust in Technology and Health Services

New technologies and data uses hold great promise, but inappropriate data use and sharing can result in people losing faith and trust in promising technologies. Long-standing harmful practices have resulted in distrust and skepticism for many communities.

First, trust can be diminished when entities that collect, share, and use consumer health information target data that is not necessary to the core product or function that people are looking for. For example, some makers of popular at-home DNA test kits not only collect sensitive DNA data, but also other user health data that may not be necessary to provide the DNA analysis services that consumers have purchased.⁴⁸ Moreover, while “these companies do a relatively decent job of protecting your DNA data,”⁴⁹ several types of non-DNA data gathered by these companies are not treated the same way and may be overshared with third parties.⁵⁰ Data collection and sharing practices that are too broad and expansive can harm people and engender distrust.

Patients and consumers also can be harmed when confronted with take-it-or-leave-it propositions regarding how health data will be collected, processed, shared, and retained. In order to access and use an app, IoT device, or even secure medical treatments, users must almost always consent to using technology and the data practices associated with it without having a complete opportunity to weigh our options and make informed choices that represent how we want our health data handled. Indeed, “[i]n the internet age, it’s become repetitive and banal to simply agree to terms of service that we don’t fully understand.”⁵¹

Unethical and exploitative data practices also occur when people and communities are not told how their data will be collected or used. For example, Native Americans and Indigenous communities have been harmed by the health care system when their health data was inappropriately collected and used by medical researchers without permission.⁵² These

48 Catherine Roberts, *The Privacy Problems of Direct-to-Consumer Genetic Testing*, Consumer Reports (Jan. 11, 2022), <https://www.consumerreports.org/dna-test-kits/privacy-and-direct-to-consumer-genetic-testing-dna-test-kits-a1187212155/> [<https://perma.cc/ZXR3-PXCP>].

49 *Id.*

50 *Id.*

51 Mary Madden, *Need Medical Help? Sorry, Not Until You Sign Away Your Privacy*, MIT Technology Review (Oct. 23, 2018), <https://www.technologyreview.com/2018/10/23/66429/need-medical-help-sorry-not-until-you-sign-away-your-privacy/> [<https://perma.cc/MF2P-TPCA>].

52 Sabrina Imbler, *Training the Next Generation of Indigenous Data Scientists*, N.Y. Times (June 29, 2021), <https://www.nytimes.com/2021/06/29/science/indigenous-data-microbiome-science.html> [<https://perma.cc/Z6YR-UWJ7>].

past practices have stigmatized Native communities and failed to respect Native customs, including around those who have died.⁵³ For example, in one case, a university researcher focusing on diabetes rates in the Havasupai Tribe gave other researchers access to data without obtaining the tribe's consent.⁵⁴ When the Havasupai became aware that their data was shared without their permission, they brought a successful legal case that returned their samples and also provided additional forms of assistance.⁵⁵

53 *Id.*

54 *Id.*

55 Amy Harmon, *Indian Tribe Wins Fight to Limit Research of Its DNA*, N.Y. Times (April 21, 2010), <https://www.nytimes.com/2010/04/22/us/22dna.html> [<https://perma.cc/6AJQ-TWHS>].

The COVID-19 Pandemic Highlighted and Heightened Disparities While Also Increasing the Demand for Technology and Data

The COVID-19 pandemic has highlighted disparate health care treatments and outcomes.⁵⁶ Underlying health and social inequities put many racial and ethnic populations at increased risk of getting sick, having more severe illness, and dying from COVID-19.⁵⁷ The pandemic has also exacerbated online privacy threats.⁵⁸ COVID-19 increased demand on our traditional health care system while also creating a surge in demand for technological solutions and the data associated with them.

-
- 56 See generally Samrachana Adhikari, Nicholas P. Pantaleo, Justin M. Feldman, *Assessment of Community-Level Disparities in Coronavirus Disease 2019 (COVID-19) Infections and Deaths in Large US Metropolitan Areas*, JAMA Network Open (2020), <https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2768723> [<https://perma.cc/WK4B-E7JZ>]; See also Maria Godoy, Daniel Wood, *What Do Coronavirus Racial Disparities Look Like State By State?*, NPR (May 30, 2020), <https://www.npr.org/sections/health-shots/2020/05/30/865413079/what-do-coronavirus-racial-disparities-look-like-state-by-state> [<https://perma.cc/AV7K-CMZP>], (nationally, African-American deaths from COVID-19 are nearly two times greater than would be expected based on their share of the population); PEW Research Center, *Financial and health impacts of COVID-19 vary widely by race and ethnicity* (2020), <https://www.pewresearch.org/fact-tank/2020/05/05/financial-and-health-impacts-of-covid-19-vary-widely-by-race-and-ethnicity/> [<https://perma.cc/G2G9-2TGV>] (The coronavirus outbreak has altered life in the United States in many ways, but in key respects it has affected black and Hispanic Americans more than others.); UCLA School of Law Williams Institute, *The Impact of the Fall 2020 COVID-19 Surge on LGBT Adults in the US* (2021), <https://williamsinstitute.law.ucla.edu/wp-content/uploads/COVID-LGBT-Fall-Surge-Feb-2021.pdf> [<https://perma.cc/D5ZU-E5JP>] (many LGBT adults are at higher risk of serious illness related to COVID-19 and its resulting negative economic impacts).
- 57 See generally, Bipartisan Policy Center, *Positioning America's Public Health System for the Next Pandemic* (2021), <https://bipartisanpolicy.org/download/?file=/wp-content/uploads/2021/06/Public-Health-Report-RV2.pdf> [<https://perma.cc/AQJ9-65HC>].
- 58 See generally Niam Yaraghi, Samantha Lai, *How the Pandemic has Exacerbated Online Privacy Threats*, Brookings (Jan. 13, 2022), <https://www.brookings.edu/blog/techtank/2022/01/13/how-the-pandemic-has-exacerbated-online-privacy-threats/> [<https://perma.cc/4R28-TP3M>].

Existing inequities can be entrenched and compounded when groups that are already subject to pervasive surveillance “may be unable to opt out of medical tracking and other surveillance systems.”

The pandemic shed new light on a simmering and complex issue for many communities. The lack of access to technology continues to plague many communities. While both cell and smartphone ownership rates increase here in the U.S., a recent Pew study still found that gaps in ownership and access persist around factors such as age, household income, and educational attainment.⁵⁹ Without access, people cannot benefit from meaningful solutions that rely on data gathered by consumer and medical health technologies. These disparities negate the potential benefits from technological solutions like digital contact tracing and exposure notification.⁶⁰

But access to digital technologies and the associated benefits must be reconciled with the data practices associated with these technologies. As discussed above, there are a litany of harms that can come from data collection, sharing, and use practices. The inability of marginalized communities to have meaningful choice around the technologies and their associated data practices – for example to opt out of medical tracking and surveillance systems – must also be grappled with. If people are only given access to limited options that risk greater harm with little benefit or upside, communities will continue to receive unequal treatment.

For instance, a recent Social Science Research Council report notes that certain technological responses to the pandemic threaten to do additional harm to already marginalized communities.⁶¹ Specifically, the report notes that without greater actions to address physical isolation, the digital divide, and surveillance exposure, disproportionate harms to these communities will persist.⁶² Existing inequities can be entrenched and compounded

59 Pew Research Center, Mobile Fact Sheet (2021), <https://www.pewresearch.org/internet/fact-sheet/mobile/> [<https://perma.cc/Y3EJ-B7KJ>].

60 Social Science Research Council, Surveillance and the ‘New Normal’ of Covid-19: Public Health, Data, and Justice (2021), <https://covid19research.ssrc.org/public-health-surveillance-and-human-rights-network/report/> [<https://perma.cc/95KF-MBF6>].

61 *Id.*

62 *Id.*

when groups that are already subject to pervasive surveillance “may be unable to opt out of medical tracking and other surveillance systems.”⁶³

The pandemic exacerbated these inequities. When we were all directed to look to alternative solutions to traditional human interactions to slow the spread of COVID-19, those with access to technical solutions pivoted to those options, while those without struggled. Because hospitals and many healthcare clinics either closed or went remote during the pandemic, individuals turned to technology out of necessity to stay connected with their health providers and meet their health care needs. The upsurge of “telehealth” or “telemedicine” replaced in-person healthcare. This was provided often via face-to-face cameras on smartphones and computers and sometimes through basic telephone communications. Indeed, more than half of U.S. consumers attended virtual doctor visits during the pandemic.⁶⁴

The pandemic also demonstrated how data from IoT health devices can forecast where new outbreaks of COVID-19 (and other infections) are likely to occur.⁶⁵ For example, smart thermometers collect data about users’ body temperatures and then process that data to make predictions about possible disease in communities.⁶⁶ Kinsa, a company that makes smart thermometers for consumers, is expanding its data processing abilities, “building an updated data hub that will be able to take data on fevers and symptoms from its smart thermometers and app and pair that with broader health data about what’s going on in a community.”⁶⁷ This type of data collection and processing can be very helpful to public health officials, but the resulting predictions must also be used in ways that recognize that not everyone has access or the ability to use smart thermometers. They can cost more than traditional thermometers and also require users to have wireless internet. Moreover, any resulting public health decisions and allocation of resources must not be overly reliant on these types of predictions. Without more representative data, some communities may face outbreaks that do not show up on tech-influenced heat maps because they lack access to the technologies that supply data.

The pandemic has spurred new technologies, including new ways to conduct contact tracing using mobile devices. These new technologies have the potential to slow the spread of COVID-19 and other contagious pathogens. When compared to traditional contact-

63 *Id.*

64 Deloitte Center for Technology, Media & Telecommunications, How the Pandemic has Stress-Tested the Crowded Digital Home (2021), https://www2.deloitte.com/content/dam/insights/articles/6978_TMT-Connectivity-and-mobile-trends/DI_TMT-Connectivity-and-mobile-trends.pdf#page=11 [<https://perma.cc/SC43-VBZA>].

65 Bryan Walsh, *Exclusive: New Data Hub to Forecast Infectious Disease Outbreaks*, Axios (Aug. 7, 2021), <https://www.axios.com/kinsa-coronavirus-prediction-smart-thermometer-936b88c2-060e-4ec5-a366-5303614e8c12.html> [<https://perma.cc/KHG7-UG3D>].

66 *Id.*

67 *Id.*

tracing, these new applications can quickly and accurately inform a person when they were in close physical proximity to another person infected with COVID-19.

However, new tech-assisted contact tracing applications have limitations and may not be accessible to, or embraced by, everyone. Tech-assisted contact tracing apps will not help those who lack access to such technologies.⁶⁸ These new applications also represent a new form of surveillance that can be used to limit individual rights and liberties.⁶⁹ Also, for these new applications to be beneficial, they need the public to trust them with their sensitive health information.⁷⁰

As the examples above demonstrate, the pandemic has resulted in calls for more data. A recent Bipartisan Policy Center report argues that high quality data systems are needed to address and respond to outbreaks, because the response to COVID-19 in the U.S. was limited due to major gaps in data collection and reporting.⁷¹ The report observes that data regarding race, ethnicity, and other demographics has been and, in some instances, continues to be inconsistently collected.⁷² More accurate, specific, and representative data can play a key role in identifying and responding to disparities, and this data can come from both consumer-facing entities and those within the traditional health care space. Moreover, combining some of these data sets may be another way to have more complete and representative data. Without truly representative data, health authorities and providers will not have a clear understanding of everyone's health needs.

68 Social Science Research Council (2021), *supra* n. 60.

69 *Id.*

70 *Id.*

71 Bipartisan Policy Center (2021), *supra* n. 57.

72 *Id.*

Ensuring Equitable Data Practices for Health Data

Modern data use is complex, opaque, and instantaneous. Trying to determine which data sets are worthier of coverage than others is no longer feasible.

The sections above have identified inequities arising from harmful data practices in the provision of health care in the U.S. New and expanding consumer-facing health technologies can help and empower people, health care providers, and governments to achieve better health outcomes. However, the current regulatory regime fails to adequately address privacy harms and discriminatory uses of health data, particularly as the scale of health-related data expands and moves quickly and effortlessly between entities regulated by HIPAA and those that are not. We have previously published a framework for better protecting health data privacy that includes a prohibition on discriminatory use of health data.⁷³ But we must do more to prevent data practices whose associated harms can be specifically acute for underrepresented and overlooked communities.

New Definition for Health Data

U.S. laws have not kept pace with new technologies. As a result, the privacy protections associated with health data are inconsistent. This inconsistency leads to certain data uses that harm individuals, particularly those in marginalized communities.

Modern data use is complex, opaque, and instantaneous.

⁷³ Center for Democracy & Technology, Executives for Health Innovation, Proposed Consumer Privacy Framework for Health Data (2021), <https://cdt.org/insights/cdt-ehis-proposed-consumer-privacy-framework-for-health-data/> [<https://perma.cc/RUP5-UY86>].

Trying to determine which data sets are worthier of coverage than others is no longer feasible. As we consider solutions to level the playing field and ensure that all health data enjoys strong privacy protections, a logical place to start is identifying and defining the universe of health data. Once defined, we can consider what appropriate uses and protections are needed.

Data can be “health data” if it is used to make inferences about a person’s physical and/or behavioral health, even if the data appears unrelated to health on its face. It no longer makes sense to attempt to identify specific entities who must protect data about our health. Instead, the protections must be attached to the data itself. When collected, shared, or used for health purposes, this health data must be protected regardless of who collects, shares, sells, buys or uses it. The key to this approach is to focus on the nature of the information and how it is used. Indeed, this is the approach we took when defining consumer health information within our Proposed Consumer Privacy Framework for Health Data.⁷⁴

This definitional approach classifies certain data as health data even though it may not initially appear to be consumer health information. For instance, consider geolocation information. Where you live may not initially appear to be related to your physical or behavioral health but when geolocation of an individual is tracked, processed, and shared in ways that make inferences or conclusions about an individual’s health based on the places they visit, like clinics or specialists, that location data is health data.

In a similar fashion, a person’s web-browsing and online purchase histories may reveal they visited websites discussing specific diseases or support forums for those conditions. Also consider data about purchase history that shows they recently bought a knee brace, weight loss supplements, or an anti-inflammatory cream. When collected, processed, sold, or shared to make inferences about peoples’ behavioral or physical health, such web-browsing and purchase data is indeed health data.

A purpose- and use-based approach to defining consumer health data recognizes the realities of today where countless entities outside of HIPAA’s limited set of covered entities create, sell, share, and use health data. This approach also expands the universe of data to better address modern data uses. A purpose- and use-based approach also avoids the overly narrow focus on very specific medical conditions that qualify for protection.

A purpose- and use-based approach to defining health data has two additional benefits. First, it benefits people by raising the protections for all the data that is used to impact their health and wellness. Second, it creates a tech-neutral standard that will stay relevant as technology evolves. A clear understanding and approach that consistently recognizes when data is health data can go a long way to addressing data practices that fail to benefit people equitably.

⁷⁴ *Id.*

Social Determinants of Health Data is Consumer Health Information

This paper has detailed how data from apps, services, and platforms outside of traditional HIPAA covered entities frequently finds its way back into the health system and influences decisions made by health care providers. An example of this type of data is frequently referred to as Social Determinants of Health (SDOH) data.⁷⁵ This is a rich data set that includes information about a person's affordable and safe housing, education, sustained income, food security, educational attainment, and living environments, which are integral contributors to health and well-being. More specifically, “[s]ocial determinants of health are conditions in the environments in which people are born, live, learn, work, play, worship, and age that affect a wide range of health, functioning, and quality-of-life outcomes and risks.”⁷⁶

Accurate and truly representative research informed by SDOH data may help address long-standing inequities.⁷⁷ It is critical to remember that these data sets are based on information about economic, societal, and environmental factors, and not individual health behaviors. Understanding this data set may place entities in a better place to address the root causes of inequity. Indeed, “[i]mproving the conditions in which we live, learn, work, and play and the quality of our relationships will create a healthier population, society, and workforce.”⁷⁸

Health care providers are looking for greater access to these data sets to have a more complete picture of patients. Today, a key “challenge to meaningfully address SDOH is the fragmented communication and coordination between the public and private sectors providing clinical, social, and human services, and with the individuals and communities served.”⁷⁹ This lack of sharing and fragmentation contribute to “unfavorable consequences, including limiting the effectiveness of resource availability and allocation, negatively impacting the quality of care, and damaging health outcomes.”⁸⁰

Some communities “may also have the greatest discomfort – justifiably – with their SDOH information being collected and shared because they are more likely to have experienced disparate or biased health care previously, are at greatest risk for misuse of SDOH data, and may face greater challenges accessing digital solutions.”⁸¹ In the same vein, “[s]ome

75 This data can also be referred to as “social drivers of health.”

76 Office of Disease Prevention and Health Promotion, Social Determinants of Health (2020), <https://www.healthypeople.gov/2020/topics-objectives/topic/social-determinants-of-health> [<https://perma.cc/373N-JJLD>].

77 National Alliance to Impact the Social Determinants of Health, Social Determinants of Health Data Interoperability (2020), https://www.nasdo.org/wp-content/uploads/2020/08/NASDOH-Data-Interoperability_FINAL.pdf [<https://perma.cc/PFK6-744E>].

78 Office of Disease Prevention and Health Promotion (2020), *supra* n. 76.

79 National Alliance to Impact the Social Determinants of Health (2020), *supra* n. 77.

80 *Id.*

81 *Id.*

individuals do not have the capacity to bear the responsibility for coordinating their data sharing, and we must avoid solutions which unfairly and inappropriately shift the expectation and burden for information sharing to individuals who may not have the resources or desire to do so.”⁸²

SDOH data must be used appropriately. A 2019 *Annals of Family Medicine* report critically examines the risks associated with certain uses of SDOH data and how some efforts could actually worsen a person’s health and widen health inequities.⁸³ Specifically, the report notes how certain SDOH uses in Medicaid, informed risk prediction models, and advances in precision medicine may be harmful and lead to reductions in the quality or access to care for certain groups of patients.⁸⁴

In an effort to use health data in a manner that reduces potential harms, equity considerations should be included in every part of the data ecosystem, including SDOH data – from product concept and design through implementation and release, and continue when processing, retaining, and sharing consumer health data. Any proposed solutions should be sure to be “sensitive to the concerns and circumstances of individuals, and technical solutions must accommodate individuals’ needs and preferences.”⁸⁵

Finally, it is important to distinguish between SDOH data and data regarding a person’s lifestyle choices. The realities that SDOH data captures are not always driven by personal choices. For example, air pollution can result in respiratory health issues, but people have no control over the air quality where they live. Data measuring air quality is SDOH data and may be used to better address community health needs.

Appropriate Collection and Use of Data About Race or Sexual Orientation May Help Address Long-Standing Inequities

Data that reflects peoples’ sexual orientation, when shared appropriately and with an individual’s knowledge and consent, can be an important information for health care providers. A *New York Times* article details how technological solutions are being developed to utilize, “the wealth of personal data in electronic health records to identify patients at high risk of getting infected with H.I.V.”⁸⁶ The algorithm at the heart of the article can be used by health care providers to better identify at-risk “patients and then steer them to a daily pill to

82 *Id.*

83 *Annals of Family Medicine, Integrating Social and Medical Care: Could it Worsen Health and Increase Inequity?* (2019), <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6342587/> [<https://perma.cc/G52Z-AMXD>].

84 *Id.*

85 National Alliance to Impact the Social Determinants of Health (2020), *supra* n. 77.

86 Gina Kolata, *Would You Want a Computer to Judge Your Risk of H.I.V. Infection?*, *N.Y. Times* (July 30, 2019), <https://www.nytimes.com/2019/07/30/health/hiv-aids-prep.html> [<https://perma.cc/Q3SV-N59F>].

prevent infection, a strategy known as PrEP.”⁸⁷ However, as with any artificial intelligence or machine learning (AI/ML) system, the predictions and models produced will be most beneficial when they are predicated on accurate and representative data.

For certain communities, past and ongoing harms have made it more difficult to establish the trust necessary to share personal information. Patients are not always keen to share this information with their doctors. One reason is that these interactions are not always positive – for instance, doctors do not always ask about or discuss sexual orientation in an appropriate manner.

Attitudes towards specific communities can be biased, paternalistic, and offensive. For example, the same article recalls a conversation when a person “asked his doctor to prescribe PrEP” and the doctor replied: ‘Why don’t you just stop having so much sex?’⁸⁸ This reported interaction is an example of an all-too-frequent occurrence for people from some communities. In this case, there is a real and viable medical solution but the doctor instead misdirects and attempts to burden the patient.

This example, and others like the racial disparities in access to kidney transplants discussed above, show that data collected about peoples’ health can be powerful when it comes to addressing the needs of underrepresented and overlooked communities’ health outcomes. But that data, regardless of whether its collection occurs within or outside of the traditional health care system, must be used in appropriate and equitable ways. Data should not be processed and shared in ways that artificially erect barriers to real treatments for underserved communities. Consumer health technologies must provide data that allows health care professionals and people to work together to solve health care challenges. To address this, clear prohibitions around discriminatory data uses against marginalized communities must be in place.

Appropriate Collection and Use of Data that Reflects Disability May Help Address Long-Standing Inequities

There are inherent tensions when addressing whether disability status should always be treated as health information. When designing, contemplating, and implementing adequate anti-discrimination protections for those in the disability community into consumer health tech, it is important to recognize the autonomy of each person and not place artificial barriers where they are not necessary.

Take, for example, a consumer health app designed to remotely monitor certain health metrics of a person, like blood pressure. On its face, this technology may be cheaper to deploy and operate compared to in-person alternatives. However, if that data ends up being used to provide, influence, modify, or deny services to someone with a disability, that app or service may be flawed.

⁸⁷ *Id.*

⁸⁸ *Id.*

Moreover, when data collected by a consumer health app or technical service limits the options available to a person with a disability, it can unfairly deny that person the freedom and ability to make their own meaningful decisions about their health care. All too often, people with disabilities do not always have the freedom to take risks because our current health care, long-term care, and benefits systems can be overly paternalistic and protective. The removal of choice can result in people being discriminated against and artificially denied their right to autonomy – they no longer have the dignity of risk. People with disabilities must have the same right to take risks as anyone else.

As such, it is necessary to grapple with hard questions around a person's disability status. To the greatest extent possible, data about disabilities should not be processed or shared in a manner that removes individual choice. Additionally, people from underrepresented and overlooked communities, including those with disabilities, interact and share their health information with others in order to access community-based services. These interactions may create situations and attitudes where certain people feel they must share data in order to simply gain access to their communities. Data collection, sharing and use approaches must consider if and when disability data about a person is being used for health purposes, as discussed in the section above, and also if there are ways to increase the protections around a person's consumer health data to ensure that this data is not used for secondary or unintended purposes that limit individual autonomy.

Better Research Practices Using Consumer Health Information

Responsible research allows for consumer health information to be collected and used in the public interest, for scientific or historical research purposes, or for statistical purposes that adhere to commonly accepted ethical standards and laws, including data from traditionally underrepresented and marginalized groups.⁸⁹ However, we should create data collection, sharing, and use practices for research purposes that avoid entrenching existing disparities, biases and harms.

To start, we can look to long-standing laws and approaches that attempt to balance between individual privacy interests, societal benefits from the use of this data, and participants' needs to process data to deliver the service or product requested by a person. However, that is only the first step. Many current approaches—including HIPAA and FTC enforcement against unfair and deceptive trade practices—do not fully address disparities, biases, and harms.

Engaging communities directly is important for any research regime. Tapping directly into community knowledge and expertise to better understand and address its needs can lead to better solutions. Partnerships of trust and engagement are essential. When communities not only know how their data may be used, shared, and retained, but are involved in shaping the data collecting, sharing, and use, they can trust others with their information and the potential for better and more representative data increases. When communities are engaged this way, they recognize the tangible returns and benefits that come from data sharing.

⁸⁹ Center for Democracy & Technology (2021), *supra* n. 73.

But building trust can be challenging in the wake of a history with inequitable and racist health studies and practices, including those discussed in earlier sections. Another way to address long-standing mistrust between researchers and communities is to ensure greater inclusion in all aspects of research so that the resulting technologies and treatments address the needs of all communities. Inclusion means transparent and participatory practices that allow all communities access to understand how their data will be used and, in turn, permit truly informed decisions regarding the use of their health data. This can even include community level review boards. Inclusion also means the participation of people from all communities to lead and inform research strategy and efforts and analyze its results.

The benefits of research using consumer health information should be relayed and communicated by study sponsors not just to people whose data will be collected and used, but to trusted members and leaders within communities. Researchers should also allow communities to help set the questions, so that research is responsive and directed toward issues that are of concern to the community and not just what the researchers believe is important. Researchers frequently rely on trusted people within underserved communities, like pastors and non-profit leaders, to reach community members. Not only can these trusted members of communities help relay and demonstrate the beneficial outcomes from research, but they can also be key advocates for their communities on the front end when they interact with researchers.


Frequently, people feel that data about their health is already being shared and they have no control over it. Empowering people, including through trusted community members, may help address bias issues on the front end – both in data collection and identification of bias within data sets and outcomes. To better ensure that resulting research and data uses are consistent with community expectations and benefit data subjects, we must create equitable standards for all research. Community-based standards, when correctly designed and followed, can help build trust, ensure that community voices are represented, and yield beneficial outcomes.

Moving forward, more must be done to ensure that research utilizing consumer health information collected from populations that have been traditionally underrepresented in research is transparent and benefits the communities where the data originates.

Also, to better address inequitable distributions of the benefits that research yields, research proposals should also include greater transparency and participation requirements that inform data subjects about the outcomes of research that use their data and address stated needs of the community.

Our work has also revealed a pronounced imbalance regarding who benefits and who is harmed by certain approaches to health research. Specifically, more must be done to move the burden of data management off consumers and onto entities that collect, process, retain, and share consumer health data. Examples of this include clear secondary use prohibitions and increased transparency provisions.

Our work has revealed a pronounced imbalance regarding who benefits and who is harmed by certain approaches to health research.



Those seeking to collect consumer health information for research purposes must inform potential research participants about how their data will be collected and used and how the research will benefit participants. Additionally, encouraging researchers to incorporate research design elements that align with equity principles can help ensure subsequent research is beneficial to people from all backgrounds.

Finally, it may be appropriate to include an increased duty of care with collection and stewardship of the research data coming from and directed at marginalized communities. Such a provision would be designed to place more responsibility onto researchers.

Diversity of Data

There are major equity implications around how data is collected and categorized. For example, when data-collecting regimes are being developed, designers have a lot of flexibility when it comes to determining exactly what data their consumer-facing tech will collect and how that data will be classified.

For example, if entities plan to collect data about race, the options presented to the user will directly impact how rich and diverse the resulting data set will be. Or, when a consumer health product collects data about sexual orientation, how many options does a consumer have to identify themselves? Are people only presented with outdated binary choices where certain people may not identify, or are there more nuanced options that allow people, should they wish, to more accurately identify themselves?

When consumer health information is being used for research or decisions that affect our health, overly simplistic datasets can result in people being underrepresented or not represented at all due to data collection and categorization choices.

Diversity of consumer health information for certain uses like research can have considerable equity implications. Population data and data used to train AI and ML systems,

like those discussed above, must do better to ensure that data accurately reflects the diversity of the people who will be affected.

Moreover, consumer health data practices need to break down artificial boundaries. Researchers using health data must work with communities, civil rights organizations, and sociologists to better explain, engage, and inform communities about the ways health data is collected and categorized, and encourage entities that are collecting and using consumer health information to be “intentional about the names of categories and vet language and displays with community members to ensure fidelity with how people self-identify.”⁹⁰

Finally, entities collecting consumer health data should actively consider and include data diversity and equity issues throughout the entire lifecycle of their consumer health offerings, covering key questions like, “What are the intended data uses?” and, “How can their offerings address the nuanced needs of communities?”

More nuanced and detailed data can also increase the likelihood that people can be personally identified. Indeed, research indicates it is difficult to truly de-identify and aggregate certain data.⁹¹ The challenges of de-identification can be especially important to consider for certain communities. For instance, it may be easier to re-identify a specific person with an uncommon condition, i.e. data points that occur less frequently. Additionally, because of the unique presentations of certain disabilities, aggregated data may misrepresent or overgeneralize and therefore not be as widely useful for disability-related studies. Therefore, certain unique data points or combinations of data points may limit how well data de-identification or aggregation can be performed.

Properly aggregated data may pose fewer privacy risks to people, families, and communities. But aggregation and de-identification are not silver bullets that ensure individual privacy. Entities must safeguard aggregated and de-identified health data from reidentification and contractually require the same commitment from any entity that receives the aggregated data. Moving forward, we must strike an appropriate balance between protecting individual privacy and allowing for diverse data to be collected and used in truly beneficial ways.

90 PolicyLink, Powering Health Equity Action with Online Data Tools:10 Design Principles (2017), <https://nationalequityatlas.org/sites/default/files/10-Design-Principles-For-Online-Data-Tools.pdf> [<https://perma.cc/78AC-FEEY>].

91 Liangyuan Na, Cong Yang, Chi-Cheng Lo, *Feasibility of Reidentifying Individuals in Large National Physical Activity Data Sets From Which Protected Health Information Has Been Removed With Use of Machine Learning*, JAMA Network Open (2018), <https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2719130> [<https://perma.cc/C9TT-SDST>].

Conclusion


When done properly, the collection, sharing, and use of individualized health information allows for dramatic improvements in physical and mental health outcomes. However, current laws and regulations do not adequately protect consumer health data against harmful data practices. We need to enact more privacy protections regarding how consumer health information is collected, processed, retained, and shared. These reforms must be rooted in fair and equitable principles to address long-standing inequities in how data practices affect the provision of health care, access to technology, and inclusion in medical research.


This paper and our Proposed Framework are made possible with the support of the Robert Wood Johnson Foundation and with assistance from our Steering Committee. Many thanks to everyone who has assisted us on this project, especially our partners at the Executives for Health Innovation (EHI).

These reforms must be rooted in fair and equitable principles to address long-standing inequities in how data practices affect the provision of health care, access to technology, and inclusion in medical research.

 cdt.org

 cdt.org/contact

 Center for Democracy & Technology
1401 K Street NW, Suite 200
Washington, D.C. 20005

 202-637-9800

 @CenDemTech