

Workshop Discussion Notes: Health

Data & Civil Rights
October 30, 2014 – Washington, D.C.
<http://www.datacivilrights.org/>

This document was produced based on notes taken during the Health workshop of the Data & Civil Rights conference. This document represents a general summary of the discussion that took place. Not all attendees were involved in every part of the conversation, nor does this document necessarily reflect the views and beliefs of individual attendees. All workshop participants received workshop materials prior to the event to spark discussion. The primer can be found at: <http://www.datacivilrights.org/pubs/2014-1030/Health.pdf>

Overview

Two distinct civil rights issues were at the forefront of the discussion of health, big data, and civil rights. The first was the issue of privacy, and the threat that greater surveillance (data collection and analysis) poses for low-income communities and communities of color. However, the sharing and analysis of big health data might improve the care of those up and down the socioeconomic spectrum, and might be particularly beneficial to low-income communities and communities of color. Analyzing health data raises a particularly vexing conundrum – the same piece of data might hold out of the hope of being used to reduce health disparities and empower people or, conversely, to violate their privacy and cause harm. The outcome depends on who holds the data and what is done with it. The area of concern also matters. Health ends up encompassing many areas, including healthcare access, medicine, and public health. Historically more independent, data has put these three into conversation in new ways. The second major civil rights issue discussed was the significant health disparities that exist within the U.S. between different populations. The promising aspect of big data in this regard would be to assist public health professionals and health care providers visualize and analyze the patterns of disease that emerge on a bigger scale, and to create interventions that address these disparities.

Themes and Discussion Topics

Are all Data Health Data?

Participants tried to define health data as information that is entered into an electronic health record (EHR) or information that is used to draw inferences about health. They debated the differences between consumer-generated and analytics data and considered differentiating observed data from inferred data. While participants acknowledged that many types of data can be health data, not all data bear on health.

Sources of health data include:

- Medical (e.g., data involved in patient care)
- Claims (e.g., insurance claims)
- Pharmaceutical (e.g., prescriptions)

- Consumer-generated
- Explicit (e.g., Fitbit, Health Kit, 23andMe)
- Inferred (e.g., Target’s analysis of purchasing data to determine pregnancy status)
- Government (e.g., Medicare, Medicaid)
- Research (e.g., clinical trials)
- Public health (e.g., infectious disease data)
- Employers (e.g., wellness programs)
- Wellness/fitness programs
- Analytics (e.g., information from data brokers)
- Marketing

Participants emphasized the need to consider the accuracy and quality of health data when used in different contexts.

- *Accuracy*: For what purpose is someone using health data? Data used for health marketing can tolerate a higher level of inaccuracy than data used for patient care.
- *Quality*: All health data is not created equal, and some health data is more valuable in different contexts. For example, data on actual health behavior is more valuable than inferred data. Medical test data is valuable in the patient care context, but it doesn’t show real-world behavior.

Challenges to Defining Uses and Harms of Health Data

Participants suggested focusing on the **uses and harms** of different types of health data rather than comparing the costs and benefits of using various types of data.

Health data is collected and used in a variety of contexts. On a broader scale, health data can support public health research and outreach or help refine analytics. On an individual level, health data can be used to help people access health care or to provide people with health care. Tension emerged around people’s expectations of how various types of data are used in health contexts, and the categories of actors that are using the data, and how and why. For example, people typically expect that prescription data is part of their health record, but they may not expect that their purchasing data can be used to make inferences about their health. People may believe that laws such as the Health Insurance Portability and Accountability Act (HIPAA) protect the privacy of their health data, but this law only protects patient care data that comes from health-care sources. Many types of health data discussed earlier are not protected under HIPAA.

It is difficult to identify harms that come from the use of health data because the same action can benefit, and harm, an individual. For example, certain population groups are underrepresented in some types of health data. To address health disparities, we can incorporate granularity into EHR systems so that a patient is labeled as “Vietnamese” instead of the more general “Asian.” However, there was some discussion that collecting more detailed information on individuals can violate their privacy or put them at risk for profiling, although the specifics of these violations were not laid out. Some populations face a double disadvantage in that there isn’t enough health data about them to improve their health outcomes, but their population is more likely to experience negative results from the use of health data beyond the realm of patient care, e.g. their data suggests a genetic predisposition to disease which could be used to deny life insurance.

Participants described a scenario that illustrated the complexities of regulating health data.

Imagine an analysis of various data points suggests that someone has not received a flu shot. On one hand, an insurance company or a public health campaign could remind the person to get a flu shot and suggest local clinics where the person could go. Or, the inference about the person's flu shot status could be used to draw further conclusions about the individual that result in him paying a higher insurance premium. The individual can experience benefit and harm from the same piece of inferred data depending on who is making the inference and for what purpose.

Health Concerns That The Civil Rights Community Can Address

Given the vast array of health data sources and the difficulty in categorizing the uses and harms of health data analysis, participants sought to identify specific problematic uses of health data. They listed several potential areas of focus:

- Reduce health disparities and improve population health
- Improve access to health care and the quality of health care for underserved populations
- Minimize loss of privacy
- Educate people about the harms of health data that circulates within and outside of protected or confidential contexts

If people can demonstrate certain types of harm, they can use civil rights laws to seek redress in situations where their health data has been misused. From a civil rights perspective, it might make sense to focus on the health data that poses the most risk to people if it is exposed. This includes medical, claims, and government data, since these data sources tend to over-represent vulnerable populations. However, some participants mentioned that minority and low-income populations tend to be over-represented in other types of health data, such as consumer-generated health data. Participants asked whether more information is needed to know what populations are over- or under-represented in various types of health data, and they discussed whether skewed health data is a civil rights issue. For example, low-income populations are over-represented in Medicaid data, and organizations can purchase that data. This creates a disparate impact for those populations, although the specific impacts were not mapped out.

Flow of Health Data to Different Entities

Although many organizations hold health data, they do not all adhere to the same standards when it comes to protecting or using that data. As mentioned earlier, HIPAA protects patient care data comes from health-care sources. But if an individual reveals the same health information in two contexts, for example, in a doctor's office and on a social network site, the same piece of information receives two different protections. The doctor is bound by HIPAA, while the social network site is not.

Participants considered whether a tagging system could mitigate concerns around the sharing of health data among organizations. However, they quickly concluded that asking people to tag each piece of their health data with permissions would be too onerous. Participants also considered limiting health data to organizations that need it, but that surfaced the challenging question of how to define need, beyond the definitive mandates of public health interests.

While some participants highlighted the privacy concerns of sharing data, others lamented that data that should be shared is too often trapped inside EHR systems. For example, if someone from Boston travels to Washington, D.C. and ends up in the hospital, the doctor in D.C. might not be

able to access the patient’s medical records. Greater interoperability between EHR systems could benefit patients when they travel or switch health care providers. It could also help researchers examine national trends or compare different populations. Participants wondered whether policy interventions could incentivize the development of interoperable systems. For example, the government can offer incentives for “meaningful use” of data and consequently promote adherence to standards. This has typically applied to use of electronic health records (EHR), but regulations can change to broaden this mandate.

Research Using Health Data

Participants agreed that as the amount of health data grows, we need new tools that address privacy concerns while facilitating the exchange of data to promote public health research and reduce health disparities. But as the size of health databases increases, researchers will likely receive access to a dataset rather than a copy of an entire data set. This may mean that the dataset resides within an organization, and researchers can use analytics tools built on top of the data to query it. In this framework, researchers may not be able to verify individual data points, but they can pull a sample of the dataset and obtain aggregate results.

Participants also pushed for health research to investigate theories of causation that are informed by bias or discrimination. For example, historically, estrogen was thought to protect against heart disease, and heart disease in women was thus not taken seriously for a long time, and problems that may have been related to heart disease were miscategorized or attributed to other diseases. More data could help determine if theories like that are the result of discrimination, different symptoms in men and women, a lack of awareness among doctors or patients, a combination of these or additional factors, etc. While this type of research already takes place, more data and analysis could crunch much more complex data to demonstrate correlations between unexpected factors. Researchers should think about ways to harness health data to investigate the underlying causes of health outcomes so that policy makers, public health workers, civil rights advocates, and others can address health disparities.

Industry, academia, and medical professionals can also come together to address concerns around outdated data. Health care companies typically have the most up-to-date health data on individuals, while large public health datasets may be a few years old. How can these different sectors collaborate to ensure that health research uses current data while still protecting individuals’ privacy?

Areas for Further Exploration

It became clear that the participants had a lot of different, and often competing, concerns about the use of health data and the benefits and harms that emerge from different usages. Privacy came up as a theme and an issue repeatedly during the conversation. The discussion emphasized that privacy should be explored further both in the context of unprotected health data that is brokered outside of HIPAA-protected spaces, and within protected health systems, particularly with reference to creating interoperable systems. As an area for further exploration, one of the most important next steps would be to develop more of a schema on the uses/benefits/harms discussion, leaning less on the term “loss of privacy” as a harm. What are the harms, where do they come from, how can they be prevented, and the same set of questions on the benefits side. How can we categorize them and generalize to ground a discussion using common terms? Next steps should

also include a discussion with a strong health policy component. When does the potential of privacy breaches play a secondary role in conversations about improving public health? When does an emphasis on privacy become a metaphor for issues of fairness and civil rights?

More specifically, the group emphasized that to explore this area further, stakeholders need to articulate their goals, such as improving health care and the quality of health care to underserved populations. Some participants expressed the need to obtain a better articulation from the civil rights community about the definitive problems that they want solved.

Generally, the group had hopes that health data could go seamlessly to the people who need it; it was also understood that suppressing data is not the appropriate response to concerns around big data. Instead, the group hoped to develop tools to better use and augment the benefits of data, and to identify the corresponding harms expressly such that these can be suppressed.

There were several more pointed issues that emerged throughout the session, and these are outlined as a series of questions below:

1. Is HIPAA a good model of balances of protections? How could we assess this?
2. What are the implications of pools of patient data that patients volunteer in order to get better help or care?
3. If a private company wanted to provide data to a researcher, and that researcher wants to use that data in a way that is privacy-sensitive, who identifies what particular data is a red flag? Or what kinds of mechanisms, or checks and balances, can we implement to identify privacy concerns?