Risks and Rewards of Demographic Data Collection: How Effective Data Privacy Can Promote Health Equity

This four-part report series is a joint project of the National Health Law Program, DREDF, Justice in Aging, Movement Advancement Project, and Race Forward
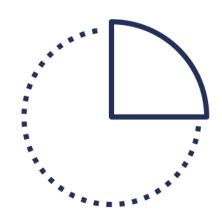
This Data, Not That Data: Big Data, Privacy, and the Impact on People with Disabilities

Silvia Yee & Mary Lou Breslin
Disability Rights Education and Defense Fund

WHY WE NEED DEMOGRAPHIC DISABILITY DATA

Disability Prevalence

Researchers and advocates working for disability data collection in health care have a saying, "Who counts depends on who is counted."[1] In other words, we cannot improve what we cannot measure. Yet health care and public health systems do not routinely collect disability demographic data even as over 64 million community-dwelling U.S. adults have a disability, representing almost 25 percent of the U.S. population.[2] The prevalence of disability is also increasing as the U.S. population ages and younger people are reporting increases in chronic conditions, making it more critical than ever for health care systems to collect and report disability data.

Health vs. Disability

Although health conditions and problems sometimes cause disability, health and disability are separate and distinct categories. That is, people can be healthy and have a disability. However, studies also consistently report health differences between people with and without disability, and those differences reflect substantial health disparities among disabled people.[3] Factors that contribute to health disparities include disadvantages stemming from disabled persons' historical segregation and exclusion from education, jobs, transportation, insurance, and housing.[4] The legacy of these practices contributes to ongoing disability stigma and helps explain why some physicians have negative attitudes toward disabled patients. It also contributes to persistent physical and procedural barriers to accessing care.[5] People with disabilities who are most affected by health and health care inequities and disparities are members of historically marginalized groups, including communities of color, LGBTQI+ individuals, and older people who have endured the intertwined effects of structural racism, ableism, lower incomes, and other forms of exclusion and discrimination.[6]

Current Status of Disability Data Collection

Federal Population and Health Surveys

Our current understanding of health disparities that affect disabled people comes primarily from federal and state population and health surveys such as the Behavioral Risk Factor Surveillance System (BRFSS) and the American Community Survey (ACS). These sources provide the most consistent data over time about how many people in the U.S. have disability and their self-reported health status. Many of these population-based surveys measure disability using six functional

limitation questions incorporated into the U.S. Department of Health and Human Services Data Collection Standards in 2011. These questions represent the minimum standard for federal population-based surveys in the U.S.[7]

Data Standard for Disability Status

Are you deaf or do you have serious difficulty hearing?
 a.   Yes   b.   No

Are you blind or do you have serious difficulty seeing, even when wearing glasses?
a.   Yes   b.   No

Because of a physical, mental, or emotional condition, do you have serious difficulty concentrating, remembering, or making decisions? (5 years old or older)
a.   Yes   b.   No

Do you have serious difficulty walking or climbing stairs? (5 years old or older)
a.   Yes   b.   No

Do you have difficulty dressing or bathing? (5 years old or older)
a.   Yes   b.   No

Because of a physical, mental, or emotional condition, do you have difficulty doing errands alone such as visiting a doctor's office or shopping? (15 years old or older)
a.   Yes   b.   No

Administrative Records

Other sources of demographic data include health care records and administrative forms that usually have some demographic data fields for race, ethnicity, and age, and sometimes for sexual orientation and gender identity, that people can mark voluntarily. However, most administrative forms and patient electronic health records (EHRs) do not have fields people with disabilities can voluntarily select to identify themselves, although they are also members of a discrete and distinct demographic group. Consequently, it is not possible to know who has a disability from these records alone.[8]

Administrative data are collected from health and health care-related documents, such as intake forms, health records, benefits applications, and needs assessments. These data are rich sources of information that help analysts compare utilization, differences in care by various patient and

community characteristics, quality outcomes, and other aspects of health care effectiveness across demographic populations. These granular administrative and health record data can provide a detailed picture of an individual's health status and changes in health conditions at the points of patient contact and at multiple times. However, without disability data fields embedded in administrative and electronic health records (EHR), health care providers, health plans, public health organizations, and state and federal health care agencies cannot identify disabled patients enrolled in health care systems as a distinct, heterogenous demographic group. This failure directly affects the quality of health care disabled people receive. It also prevents researchers from investigating the intertwined causes of health and health care inequities and disparities among people with multiple demographic identities, including disability. Moreover, the absence of this data also increases the difficulty of establishing that people with disabilities are subject to health and health care disparities caused by factors other than an individual condition or disability.

EHR Data

Over the past decade, as interoperability standards take increasing effect, the EHR has become especially important because data-sharing requirements are widely used. EHRs make patient data readily available to health care providers in multiple locations and, in some situations, allow care coordination across social service organizations and healthcare entities. These records also enable rapid assessment of public health trends and threats, such as the 2020 COVID-19 pandemic. In addition to including disability data in administrative records, it must be included and displayed in EHRs so analysts and researchers can understand and measure real-time progress toward addressing injustices and identifying pathways to improving health care quality.[9]

Practical Data Uses

More specifically, disability demographic data are required within EHRs and administrative forms to identify and provide accommodations that are fundamental to ensuring that people with disabilities receive and benefit from health care that is at least as effective as that provided to people who do not have a disability. Examples of accommodations include physically accessible buildings and medical and diagnostic equipment, extra time for a clinical visit when it is needed for effective communication, and print materials made available in accessible formats for people with vision limitations. Disability data is also needed to correctly identify health and health care disparities that disabled people experience and to analyze their impact on health status, health outcomes, future health, and longevity. It is also required to deepen our understanding of interrelated disparities experienced by disabled people who also have other characteristics associated with health disparities. And finally, disability data supplements effective civil rights enforcement that responds to discrimination experienced by people with disabilities, including those with multiple demographic characteristics.

Even as there is a powerful case for collecting disability demographic data, advocates are also concerned about the potential for widespread data misuse as artificial intelligence (AI) and algorithms incorporate vast amounts of data into health care decision-making tools that clinicians across the country use every day. The emergence of this concept of "Big Data" in health care poses increasing privacy and safety issues for disabled people, raising risks that cannot be readily mediated through setting disclosure requirements or giving individuals an opportunity to "correct the record." The following section explores some of these concerns.

THE CONSEQUENCES OF BIG DATA COLLECTION FOR PEOPLE WITH DISABILITIES

The collection of disability functional status involves walking a fine and difficult line for disabled people. On the one hand, the disability community and advocates have fought long and hard to be recognized in federal and state demographic census collection activities, health care surveillance tools, public health analysis and research, electronic health record interoperability standards, and all manner of health care quality measure development. The consequences of people with disabilities not being counted among those who are capable of good health results in a self-fulfilling prophecy that ensures disabled people can rarely achieve good health.

> **The consequences of people with disabilities not being counted among those who are capable of good health results in a self-fulfilling prophecy that ensures disabled people can rarely achieve good health.**

On the other hand, the fact that information about disabled people's spending habits, online behavior, health-related needs, relationships, reading patterns, and so forth is being swept up indiscriminately with that of the rest of the population is not going to lead to people with disabilities being recognized as a disparity population or achieving equally effective health care. In this section, we will address how Big Data differs from the kind of disability information that is collected in EHRs and by health care providers. We will examine why health-related Big Data mostly falls outside of existing privacy protections for health data. And finally, we will look at some of the negative consequences for the disability community that arise from having unregulated Big Data commercially available in the world, especially when more accurate disability demographic data is still not reliable collected by health care providers or the U.S. public health infrastructure.

How Big Data Differs from Demographic Disability Data and Why the Difference Matters

Being counted in an era of Big Data collection, combined with the increasingly common use of machine learning and algorithms that promise innovative ways to improve provider accountability, quality of care, complex diagnoses, and the efficient use of limited resources, bears striking and specific risks for disabled people. Big Data can be conceptually distinguished from "regular" data by looking at multiple dimensions that technical analysts and marketers have characterized as "V factors" ranging in number from three to nine or more.[10] For our purposes, we emphasize volume, variety, and velocity. New machine learning techniques enable very large amounts of a wide array of different kinds of data, collected from different sources, to be rapidly analyzed across many individuals and types of consumer populations.[11] These facts lead to a fourth "V" factor of variability, inconsistencies among the data, since the large amount of data collected can readily lead to discrepancies between data sources, especially since some data sources can be more quickly uploaded and updated than other sources. Big Data is *so* big that it needs special applications and processes—a Big Data architecture for storing, manipulating, analyzing, and reporting the information that is collected. Big Data is not limited to health or health care data, but such data can be incorporated within Big Data, even though the entities that generate, hold, use, or sell health-related Big Data are not regulated in the same way as health care data that is found in EHRs and provider charts.

This last fact leads to one of the biggest problems with Big Data and the impact it can have on people with disabilities and other groups whose privacy is typically protected when it comes to their health care information. The first paper in this series, *Administrative Data: Providing Information to Advance Autonomy and Drive Equality*, provided a short summary of the ways in which health care information is protected.[12] When individual patients voluntarily share demographic information with providers, hospitals, and health plans, that information cannot be further disclosed or sold because of the Health Insurance Portability and Accountability Act (HIPAA). The difficulty with Big Data, even when it involves information that we might typically characterize as related to health or health care, is that the entities that are collecting and working with Big Data are often not providers, hospitals, health plans, or even employers or schools/universities.

Big Data and the Limits of Current Privacy Protections

HIPAA was primarily motivated by privacy concerns when it was passed in 1996, but federal lawmakers also came to understand that the increasing use of electronic devices for storing patient information and the encouragement of information sharing among a patient's multiple or successive providers or insurers raised an accompanying need for updating patient privacy in these new contexts. Nonetheless, the "Privacy Rule" regulation was not promulgated by the Department of

Health and Human Services (HHS) until 2000, coming into effect in 2003 (with an additional one year extension for small health plans).[13]

The Privacy Rule does not broadly apply to any kind of entity with health-related information, and it does not apply to all types of health information. Instead, HIPAA's privacy protections are triggered when the entity that is collecting or generating the information is a covered entity that meets the requirements in the regulation.[14] Any health plan, health care provider, or "health care clearinghouse" is a covered entity, as well as the "business associates" of these listed entities, which essentially includes the organizations and business with which they contract to handle some aspect of their health, administrative, financial, or legal data.[15] Technology companies such as Google, Meta, or Amazon, which interact with but do not directly collect or generate health-related data, may not be covered entities.

As for the information itself, HIPAA applies privacy protections to Protected Health Information (PHI), defined as "any information held by a covered entity which concerns health status, the provision of health care, or payment for health care that can be linked to an individual.[16] But the definition of Health Information is itself narrower than what the average person might think:[17]

Health information means any information, including genetic information, whether oral or recorded in any form or medium, that:

(1) Is created or received by a health care provider, health plan, public health authority, employer, life insurer, school or university, or health care clearinghouse; and

(2) 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 the past, present, or future payment for the provision of health care to an individual.

Under the above definition, the entity that creates or obtains the information is as important as whether the information concerns an individual's health care. The kind of information that is protected must relate to a specific individual, and HIPAA sets out 18 categories of information such as Social Security numbers, names, phone numbers, and email addresses that cannot be released with someone's health care information because that would allow the information to be linked to a particular person.[18]

What Can Be Done with all that Big Data?

Taking all these requirements into consideration, we can see how many collectors and sellers of Big Data health care information do not fall under HIPAA regulation. The large corporations behind social media and online search tools such as Meta and Google are not one of the listed types of health entities, and they are usually not under a contract with a health provider, plan, or insurer to work

with their data. These corporations can easily avoid disclosing sensitive information such as Social Security numbers when they work with data about what Facebook users are sharing or the kinds of searches that people are making on Amazon or Google, but that does not mean that they are collecting information that is permanently scrubbed of individual identifiers. Anyone who uses their email to sign-in to an account to expedite online purchases can have a unique identifying token attached to the email that would allow tracking of the various consumer websites that are visited with the same email, even if the email itself is not explicitly linked to the consumer information, or sold to third parties.[19] For example, someone who has a history of purchasing orthopedic equipment and mobility devices, diabetic supplies, and over-the-counter pain medications, as well as conducting online searches for personal care assistants or joining a social media group support group connected to a health condition, could have all those purchases and online interactions tracked and linked if they set up accounts to make online interactions quicker and more convenient. Online and social media companies can easily compile this data, analyze it, and impute disability status to some consumers according to what they buy and what they "like" and search online. While these companies could currently hold the relatively harmless goal of sending targeted advertisements, they are not regulated by HIPAA and are not held to storage and privacy standards for the data they hold. That means the data can be sold for profit and otherwise handled without any obligation to notify identifiable consumers.

In addition to the purchases that disabled people make, there are also numerous wearable and online apps that relate to personal health and well-being. Some are intended for general use while some are intended for use by people with disabilities. The health app industry is thriving and predicted to reach potentially $111 billion by 2025.[20] Among apps intended for general use, ones that track menstrual cycles are popular among many people who wish to become pregnant or receive more timely feedback on their own cycles. One of the most popular apps, Flo, claims to have 43 million active users while Clue, a similar app, cites 12 million active users.[21] Some reproductive rights and health advocates have noted that these apps raise a risk of prosecution in states that have taken, or plan to take, steps to criminalize abortion, because the health information collected can potentially reveal when a pregnancy started and stopped. If the app's stored information were subpoenaed or sold (and investigative journalists have already found that Flo sold information to Facebook about when users were having their period or planned to get pregnant), it could be combined with additional Big Data information to discern identity at the individual level. And note again, that most commercial social media and health and fitness app designers deal primarily in data, even if it is data about health. They do not provide or pay for health care, even when they offer apps that can be particularly helpful to people with disabilities or people with health conditions, and therefore are likely not regulated by HIPAA requirements. As a result, the data derived from apps marketed specifically to disabled people is fair game for Big Data collection.

In cautioning against the use of menstrual cycle apps in states where abortions are being criminalized, one advocate noted that "the most secure option might just be the most old-fashioned: tracking your cycle on paper."[22] The suggestion, however, is deeply problematic for some people who have disabilities that affect their vision, mobility, or ability to write and read easily. The bright promise of technology for many people with disabilities is that it enables them to more easily and privately perform necessary tasks and independently maintain information in their lives. Many companies explicitly address this need when they offer apps that help track and manage chronic pain, provide task reminders, transcribe the speech of an incoming caller for deaf or hard of hearing persons, provide wheelchair accessibility information for an urban location, or read aloud results from a diabetic or other monitoring device for people who are blind or low vision. In addition to being free or relatively inexpensive, many of these apps will sync across multiple devices and allow for the sharing of lists and other information, so that a disabled person can, for example, share their blood sugar levels with their spouse or parents.[23] The suggestion that disabled people must make some kind of devil's bargain and agree to give up privacy protections when they use wearable or online apps that help them to self-monitor their health or seek to share that information with family or their health providers is not acceptable. Equally unacceptable is holding disabled individuals responsible for scrutinizing a needed app's disclosures and constantly monitoring its data policies for unwanted changes.

WHY HEALTH-RELATED BIG DATA IS A PARTICULAR RISK FOR PEOPLE WITH DISABILITIES

Even beyond the pure loss of privacy implications for people who have a disability-related need for health-related mobile apps and wearable technology, Big Data presents deeper problems for the disability community because of how far behind health, health care, and public health systems are when it comes to collecting demographic disability and functional information. If Big Data on disabled people is collected and superficially analyzed, then it can also become available for HIPAA-covered entities to acquire. Once acquired by covered entities, the information would likely be subject to HIPAA protections; however, there is nothing in HIPAA that disallows health providers, plans and hospitals from using that information for their own analyses and diagnoses, treatment, and resource allocation decisions among patients, even if they cannot further "release" the data.

Some Consequences of Using the Wrong Data

While the incorporation of Big Data by health providers might lead to catching some missing disability data in health care, it is much less preferable than allowing people to self-identify their disability functional status. Including Big Data components into health care machine learning and algorithms without context and any ability to distinguish between correlation and causation or to discern how disability discrimination, stereotype and prejudice figure in that information, has great potential to harm people with disabilities more than it will help fill in gaps in data. The conflation of correlation

and causation, that is, the assumption that two things happening at the same time must mean that one causes the other, has always been deeply problematic for disabled people in the health arena. For example, some people with various disabilities live in skilled nursing homes. The general public has historically viewed this as something that naturally happens when someone becomes significantly disabled because disability causes the individual to lose their health and capacity to take care of themselves and live in the community. But disability advocates would argue that the correlation between disability and nursing homes occurs because the kinds of community services and supports that would allow people with significant disabilities to live in the community are underfunded and not always available, accessible and affordable. Housing is hard to find, and some persons who newly lose functional capacities like walking are socialized to expect that they will end their days in a nursing home. If Big Data links disabled people and their families with wheelchair purchases, online searches for nursing home options, and social media campaigns for nursing home safety, it reinforces the idea that disability triggers a need for nursing homes, even though the truth could be that living in nursing homes causes increased physical and mental incapacity over time.

The risks of misusing Big Data to make incorrect assumptions and false correlations have been documented in real life. For example, reporting has uncovered how an algorithm used by many doctors and pharmacists to determine individuals who are at higher risk of opioid addition is likely sweeping in a wide range of certain health indicators (*i.e.*, presence of chronic pain, experience of mental health trauma or abuse, relationships with multiple providers, use of certain pain medications) to flag disabled persons as likely candidates for opioid addiction. The predictive scores are based on factors that co-exist among individuals who have become addicted to opioids and there is no further investigation into whether the correspondences can be factually used to predict behavior. There is also no consideration of how some of the factors, such as a relationship with multiple providers, are reasonably expected for individuals with complex disabilities and chronic health conditions. Nor is there any consideration for how a denial of needed, effective pain medications could leave a disabled person **more** likely to seek illegal drugs, fewer options for living a healthy and independent life, and more likely to consider suicide given barriers to obtaining effective pain relief.[24]

An example of how wrong data can be used to reach an incorrect conclusion can be found in algorithms used by state Medicaid agencies to determine resource allocation for home and community-based services. Such algorithms commonly fail to consider demographic disability data that the first part of this brief identifies as necessary, in large part because states fail to collect this information, and do not make any allowance for barriers to consistently obtaining accommodations and policy modifications that disabled Medicaid beneficiaries need for effective care. Instead, a Medicaid agency could choose to incorporate Big Data, and assume that any individual who manages to be out in the community – as shown by fitness trackers or shopping or reminder apps – must be doing physically well and therefore can get by with reduced service hours. We could see a pernicious interaction of algorithms using the **wrong** data to make treatment and resource allocation decisions. The ubiquitous lack of functional disability status questions in state and other health care surveys,

combined with shallow, stereotypical, or just incorrect assumptions about the health of people with disabilities and the level of function needed to live safely in the community, will lead to biased algorithms and treatment recommendations.

Deadly Consequences for Disabled People

The potentially tragic juxtaposition of Big Data being readily available when disability demographic data is not is illustrated in an article published in 2019 that focused on how Big Data raises "big risks and challenges, among them significant questions about patient privacy."[25] The authors recognize the privacy concerns brought up by the use of Big Data but also posit a cost to patients in forgoing Big Data use, supplying the following vignette as a way to illustrate the nuance of evaluating where a privacy breach has occurred and the risks of a foregone benefit.

> Samantha presents at Chicago Hope Hospital with moderate organ dysfunction; the physician is trying to decide whether to send Samantha to a specialized ICU; Samantha might benefit, but beds are limited and other patients might benefit more. In traditional medicine, assessing a patient's risk for cardiopulmonary arrest or other preventable serious adverse events might take hours, the assessment has limited prognostic accuracy, and the risk may change during that period. Imagine that an alternative assessment mechanism is available. CorazonAI has developed a predictive analytics engine, based on data from millions of U.S. patients' electronic health records, that could ascertain the risk accurately for hundreds of patients with real time updates to help the physician evaluate who should be admitted to the ICU. The physician uses the system, which advises that Samantha be admitted.

In this vignette, Big Data is the hero, providing a more objective and evidence-based way to distinguish patients who will most benefit from scarce hospital resources. Strikingly, this scenario of scarce hospital resources eerily forecasted a horrifying reality with deep resonance for people with disabilities that occurred during the COVID-19 pandemic, when hospitals throughout the country were overrun with patients in profound respiratory distress who needed breathing assistance and specially trained personnel that were only available in intensive care units (ICUs). Though COVID-19 infection spread widely in the U.S., its health impact was not equal among racial groups,[26] and it was also not equal for disabled people and older persons who lived in congregate care facilities.[27] For people with disabilities who live in the community, which is most people with disabilities, the lack of functional limitation information in EHRs for people with vision, speech, communication, intellectual, and other disabilities affected how emergency department providers would rank them for ICU treatment using medical severity or morbidity tests in widespread use, such as the Glasgow Coma Score (GCS)[28] or the Sequential Organ Failure Assessment (SOFA) Score.[29] The GCS measures visual, verbal, and motor responses in patients with traumatic brain injury, in a coma, or in critical condition. People with vision, hearing, communication, or dexterity disabilities might not be responsive to spoken commands to move a certain way or articulate speech, not necessarily because of the

Providers use these common predictive medical tools to make life and death decisions on who gets prioritized for treatment, but the tool itself is usually applied in ways that incorporate incorrect assumptions about how responsive patients act.

severity of a coma or brain injury, but because they have underlying functional limitations that affect their capacity to respond to the external stimulus use the CGS.[30] Similarly, individuals with significant mobility disabilities or other chronic conditions may have respiratory limitations that require additional time to become fully responsive to ventilator treatment provided in an ICU, and if they are denied that needed accommodation they will be consistently evaluated with a SOFA score already weighted toward being unable to benefit from treatment.[31] Providers use these common predictive medical tools to make life and death decisions on who gets prioritized for treatment, but the tool itself is usually applied in ways that incorporate incorrect assumptions about how responsive patients act. The tools fail to take account of how extending accommodations for individual functional limitations would give a more accurate assessment of how specific individuals with disabilities would respond to treatment.[32] Since EHRs fail to track functional limitation and accommodation information, people with disabilities were left terrified of winding up in an emergency department and unable to indicate their functional capacity or needs fully, leaving them likely to be deprioritized for treatment should care be rationed during a surge situation.[33]

Clearly, an article published in early 2019 could not be prescient about the degree to which morbidity tools would fail disabled people during a worldwide pandemic that began almost a year later. But comparing rosy predictions about the universal helpfulness of predictive analytics tools that use "millions of U.S. patients' electronic health records" with the reality that those records *do not contain vital information about disability demographic status* is a helpful reality check. When the privacy implications of Big Data analytic tools are examined from the stance of legal, ethical, or best practices, the situation of patients with disabilities is rarely considered as something that merits special consideration, or any consideration at all.[34]

The fact that disability status is not commonly collected may give the impression that any consideration of privacy issues is premature; the privacy of disabled persons should be protected once we ask them to volunteer information about their disabilities. To the contrary, health-related Big Data makes it incredibly easy to infer who has a disability with consequent concerns for how that information can be used, discovered, and corrected, even if it is not further shared. Moreover, as health systems and governments take steps to address inequities in health care quality and access to social drivers of health, adding disability functional status to race, ethnicity, language, sexual

orientation, and gender identity, will also lead to greater potential for reidentification. The privacy needs of patients with disabilities need to be understood and addressed now and not only some time in the future when formal collection of disability demographic information has begun.

COLLECTING THE "RIGHT" DISABILITY DATA AND PROTECTING THE PRIVACY OF ALL DATA

The disability community, in all its rich heterogeneity of race, ethnicity, age, sex, sexual orientation, and gender identity, does not face a true dichotomy between collecting functional disability status information or health-related Big Data. There is really no way to insulate disabled individuals from an engagement with Big Data collection activities, and too much to gain from the convenient online health apps and personal devices that contribute to the amassing of Big Data to try and cloak people with disabilities from Big Data archives.

The proliferation of wearable devices and health monitoring apps, corporate ingenuity for turning a profit, and the great usefulness of being able to quickly share health data that others need, whether those others are family members or health care providers, will result, and has already resulted, in releasing great swathes of health-related information out in the world. By choice, for convenience, and for health care and coordination reasons, health care Big Data is already being collected, analyzed and sold. Even when it is done fraudulently by entities that are caught and punished, there is no realistic way to put the information genie back in a bottle, especially after several years of release.[35] It will be used to feed health care algorithms that are already prone to incorporate normalized values that place a lower value on health care interventions that do not "fix" disability and cure chronic conditions.[36] And as a result, some of this data is likely to show that at least some people with a wide range of disabilities experience worse outcomes, shorter lives, poorer health, slower recovery, and higher health care expenses over a longer period of time than non-disabled persons. In short, Big Data collection and use is likely to confirm the assumptions of many health care professionals, academics, and even some ethicists that disability equates with poor health and a less worthy life. Without the collection of functional disability status and a true analysis of health and health care disparities experienced by people with disabilities, disabled people and advocates will be unable to establish the impact of stereotype, inaccessibility, discrimination, and implicit bias on health outcomes for disabled persons.

Practically speaking, there are at least two scenarios that raise the urgent need to develop privacy protections for health-related Big Data. One scenario recognizes that the unique identifiers often attached to Big Data will leave people with disabilities identifiable to entities who could find or purchase health-related information from Big Data entities that are not subject to data controls, not even a minimum requirement to disclose that data is being collected and how it could be used. As a result, health-related data can be freely used to infer the presence of disability with discriminatory consequences. For example, a landlord seeking tenants could use Big Data by way of employing an

algorithm that, unbeknownst to consumers, discriminatorily screens out people who potentially have a mental health disability or families who have disabled children. There is already evidence that Black women are disproportionately represented in eviction court proceedings, and therefore algorithmic tenant screening tools that use eviction records without any context are likely to incorporate this disproportionality.[37] A Black woman who has been evicted in the past, who is also discoverable through Big Data tools as someone who likely has disabled children because she engages in special education advocacy through social media may quickly sink to the bottom of many landlords' list of desired tenants simply due to the intersection of multiple stereotypes.

Even if Big Data is used or sold in a de-identified form, it could shape the general assumptions and links made between health conditions and outcomes that will be used in analytic diagnostic, treatment, and resource allocation tools, even as individualized patient information about functional disability fails to be collected and is not available to repudiate or modify those generic assumptions. This is already happening among health insurers who may not be using Big Data yet, but instead are using algorithms based on generalized patient health care data to make predictions based on factors such as "diagnosis, age, living situation, and physical function" to assess an individual Medicare beneficiary's mobility, cognitive capacity, and medical need. The algorithm proposes a length of stay and a target discharge date that also acts as the cut off point for the insurer's coverage of skilled nursing home payments, even though the target cut-off is contradicted by every treating provider.[38] If algorithms and AI are not required to include individualized accurate disability demographic data, it is very easy to make the leap from using general patient record data to using Big Data that is readily available and not subject to privacy protections and nondiscrimination requirements.

As a final example, consider an individual with a medical diagnosis of Cystic Fibrosis. The diagnosis can indicate limited function in a person's lungs or other organs, but this is not necessarily the case for everyone who has the condition. An analytic tool for finding prognosis upon hospitalization may provide an expected health outcome that fails to make an individualized assessment for a specific patient's historically recorded breathing capacity, overall health, and accommodation needs.[39] At the same time, if this individualized patient information was readily available in the EHR, and considered on par with an analytics tool that uses Big Data, the combined impact might help obviate against the subjective opinion and implicit bias of providers who have also been trained to think of a Cystic Fibrosis diagnoses as indicative of a shortened life span and lower quality of life. In short, health-related Big Data could help people with disabilities by mediating against provider explicit and implicit bias, but only if that data is considered in conjunction with individualized patient information on functional limitation and accommodation needs, especially since individualized EHR data is far more likely to be viewed and corrected by patients themselves and therefore more accurate.

CONCLUSION AND RECOMMENDATIONS

Ultimately, health-related Big Data uniquely implicates ill health for people with disabilities because of the direct link that is often erroneously drawn between disability and lack of health, though age can also be treated as a shortcut proxy for disability.[40] Once individuals are identifiably associated with disability, they will be subject to multiple risks of discriminatory consequences. And once disabled people experience increased discrimination and/or stigma because of being identified with having a disability, they will be even less likely to voluntarily disclose their functional disability status. This chain of consequences is exacerbated by the unregulated nature of health-related Big Data collection, use, and sale and its already common dissemination. Nonetheless, the substantial benefits of

The substantial benefits of collecting voluntarily disclosed demographic disability data will help offset the risks of Big Data if disabled people's privacy is better protected.

collecting voluntarily disclosed demographic disability data can help outweigh disability data-collection risks for all the reasons provided at the beginning of this brief, especially if the abuse of health-related Big Data is reined in. For people with disabilities, privacy protections need to be extended and they must be extended now, in conjunction with an education campaign that informs disabled people about the importance of disclosing disability in HIPAA-protected contexts, and amidst renewed efforts to consistently include functional disability status in EHRs and in public and private health programs and activities.

Recommendations

Require health care entities to collect voluntarily-provided functional disability status as a part of demographic information collection in electronic health records and program applications, and develop and adopt uniform national standards for such data collection.

Fund and incentivize research on the best ways to solicit accurate answers to demographic data collection questions on disability, including a review of whether additional questions are needed to close gaps in granular disability information that lead to undercounting individuals with certain disabilities.

Fully incorporate disability into research on how health-related Big Data can lead to discriminatory consequences and inferences for groups identifiable by race, ethnicity, preferred language, age, sexual orientation, gender identity, variations in sex characteristics, and so forth.

Develop privacy protections for health care information that apply appropriately to entities that collect, store, analyze, and sell health-related Big Data but are not covered by HIPAA.[41]

Ensure that existing non-discrimination protections, such as Section 504 of the Rehabilitation Act and Section 1557 of the Affordable Care Act, apply to the ill consequences that can arise if Big Data entities, or those to whom they sell Big Data or data analyses, indiscriminately or erroneously collect, store, analyze, or sell, health-related information.

Hold insurers and other HIPAA-covered health entities accountable for their use of Big Data by making them responsible for outreach and public education on Big Data and how it can be distinguished from demographic information in EHRs, requiring transparency in their uses of Big Data, and holding them accountable for any leaks of health-related Big Data. **Notification is not enough in itself**.[42]

ENDNOTES

[1] N. S. Reed, *et al.*, "Disability and COVID-19: Who Counts Depends on Who Is Counted," *Lancet Public Health*, vol. 5, no. 8 (2020): e423.

[2] "Community-dwelling" refers to adults living at home as opposed to in an institution. Centers for Disease Control and Prevention (CDC), National Center on Birth Defects and Developmental Disabilities, Division of Human Development and Disability, Disability and Health Data System (DHDS) Data (2020) [accessed Feb 12, 2023].

[3] G. L. Krahn, *et al.*, "Persons with Disabilities as an Unrecognized Health Disparity Population," *American Journal of Public Health*, (Apr. 1, 2015), 105:S2, S198-S206; CDC, *Disability and Health Data System (DHDS) Data* (2020).

[4] L. I. Iezzoni, *et al.*, "Have Almost Fifty Years of Disability Civil Rights Laws Achieved Equitable Care?" *Health Affairs (Millwood)*, (Oct. 2022), 41:10, 1371-78; W. Horner-Johnson, "Disability, Intersectionality, and Inequity: Life at the Margins," in D. J. Lollar, W. Horner-Johnson, K. Froehlich-Grobe, eds., *Public Health Perspectives on Disability: Science, Social Justice, Ethics, and Beyond*, New York (NY): Springer US, (2021), p. 91–105.

[5] T. Lagu, C. Haywood, K. Reimold, *et al.*, "'I Am Not the Doctor for You': Physicians' Attitudes About Caring for People with Disabilities," *Health Affairs (Millwood)*, (Oct. 2022), 41:10, 1387-95

[6] S. Yee, M. L. Breslin, T. Goode, *et al.*, "Compounded Disparities: Health Equity at the Intersection of Disability, Race, and Ethnicity," Resource White Paper, 2018, National Academies of Science, Engineering, and Medicine; National Academies of Sciences, Engineering, and Medicine, Health and Medicine Division, Board on Population Health and Public Health Practice, Roundtable on Health Literacy, Roundtable on the Promotion of Health Equity, "2. Overview," People Living with Disabilities: Health Equity, Health Disparities, and Health Literacy: Proceedings of a Workshop, Washington (D.C.), National Academies Press (U.S.), (Nov. 28, 2017).

[7] Centers for Disease Control and Prevention (CDC), "Behavioral Risk Factor Surveillance System," (last modified May 16, 2014).

[8] M. A, Morris, M. K. Hamer, K. Eberle, *et al.*, "Implementation of Collection of Patients' Disability Status by Centralized Scheduling," *The Joint Commission Journal on Quality and Patient Safety* (2021), 47:10, 627-636.

[9] M. D. Douglas, D. E. Dawes, K. B. Holden, *et al.*, "Missed Policy Opportunities to Advance Health Equity by Recording Demographic Data in Electronic Health Records," *American Journal of Public Health*, (Jul. 2015), 105,, suppl. 3, S380-8; I. Akobirshoev, M. Vetter, L. I. Iezzoni, *et al.*, "Delayed Medical Care and Unmet Care Needs Due to the Covid-19 Pandemic among Adults with Disabilities in the US," *Health Affairs (Millwood)*, (Oct. 2022), 41:10,1505-12.

[10] See, for example, Julius Černiauskas, "Understanding the 4 V's of Big Data," Forbes Technology Council Post, August 23, 2022; Matt Moore, The 7 V's of Big Data, Impact.com [Accessed March 20, 2023].

[11] National Institutes of Standards and Technology, U.S. Department of Commerce, Big Data Public Working Group, Definitions and Taxonomies Subgroup, *Big Data Interoperability Framework: Volume 1, Definitions, Version 3*, (Oct. 2019).

[12] Naomi Goldberg, "Administrative Data: Providing Information to Advance Autonomy and Drive Equality," Nat'l Health L. Prog. (Feb. 2023).

[13] Ryan Mueller (2022). "Big Data, Big Gap: Working Toward a HIPAA Framework that Covers Big Data." 97 *Indiana L. J.* 1505. "The Privacy Rule is codified in Title 45 of the Code of Federal Regulations in section 160 and subparts A and E of section 164. 45 C.F.R. §§ 160.101, 164.104, 164.500. (2019)."

[14] 45 C.F.R. § 164.500.

[15] 45 C.F.R. §§ 160.103, 164.500.

[16] HIPAA History, *HIPAA Journal*.

[17] 45 C.F.R. § 160.103 – Definitions.

[18] 45. C.F.R. § 164.514(b).

[19] Brian X. Chen, "Everyone Wants Your Email. Think Twice Before Sharing It." *New York Times* (Jan. 25, 2023).

[20] Mueller (2022) at 1513-14.

[21] Rina Torchinsky, How Period Tracking Apps and Data Privacy Fit into a Post-Roe v. Wade Climate. *NPR* (June 24, 2022).

[22] *Ibid*.

[23] See, for example "25 Life Changing Apps for People with Disabilities" (2018), or "Top 8 Mobile Apps for Persons with Disabilities" (2022).

[24] Maia Szalavitz, "The Pain Was Unbearable. So Why Did Doctors Turn Her Away?" Wired, Aug. 11, 2021.

[25] W. Nicholson Price, II & I. Glenn Cohen, "Privacy in the Age of Medical Big Data," *Nat Med*. (Jan. 7, 2019).

[26] Officer of the Assistant Secretary for Planning and Evaluation (ASPE), US Department of Health and Human Services (HHS), *Health Disparities by Race and Ethnicity during the COVID-19 Pandemic: Current Evidence and Policy Approaches*, (Mar. 15, 2021).

[27] National Council on Disability, 2021 Progress Report: The Impact of COVID-19 on People with Disabilities, Chapter 2, Impact of COVID-19 on People with Disabilities in Congregate Care Facilities, pp. 105-133, (Oct. 29, 2021).

[28] Shobhit Jain & Lindsay M. Iverson, Glasgow Coma Scale [Updated Jun. 21, 2022], StatPearls . Treasure Island (FL): StatPearls Publishing 2022 Jan. [Accessed March 16, 2023]

[29] MD+Calc, Sequential Organ Failure Assessment (SOFA) Score] [Accessed March 16, 2023].

[30] A range of opinions and disability awareness can be discerned among practicing health care providers on how CGS should be applied. See, for example, "GCS in nonverbal patients," *Online Ambulance* (May 8, 2006).

[31] Ari Ne'eman, *When It Comes to Rationing, Disability Rights Law Prohibits More than Prejudice*, Bioethics Forum Essay, The Hastings Center (Apr. 10, 2020).

[32] The federal Department of Health and Human Services' (HHS) Office for Civil Rights (OCR) published a FAQs for Healthcare Providers during the COVID-19 Public Health Emergency: Federal Civil Rights Protections for Individuals with Disabilities under Section 504 and Section 1557 (last reviewed Feb. 4, 2022) that provides a useful analysis of how and when medical tools and decision-making can discrimination against people with disabilities.

[33] Disability Rights Education and Defense Fund (DREDF), "Preventing Discrimination in the Treatment of COVID-19 Patients: The Illegality of Medical Rationing on the Basis of Disability," (Mar. 25, 2020); W. F. Hensel & L. E. Wolf, "Playing God: The Legality of Plans Denying Scarce Resources to People with Disabilities in Public Health Emergencies," *Florida L. Rev.* (2013), 63: 3-5, 720-769.

[34] See, for example, I. Glenn Cohen, Ruben Amarasingham, Anand Shah, Bin Xie, & Bernard Lo, The Legal and Ethical Concerns That Arise From Using Complex Predictive Analytics In Health Care, *Health Affairs* 33:7, 1139-1147 (July 2014). The article recommends including patients and other stakeholders in implementing governance structures for Big Data analytics tools but does not address the kinds of practical accommodations and policy modifications that are needed for the full inclusion of patients with disabilities.

[35] Federal Trade Commission, *FTC to Ban BetterHelp from Revealing Consumers' Data, Including Sensitive Mental Health Information, to Facebook and Others for Targeted Advertising*, (Mar. 2, 2023). The BetterHelp case is particularly interesting because it did not appear to involve HIPAA or any HIPAA analysis, even though the law presumably applied to BetterHelp as a mental health care provider.

[36] Ian Moura, Encoding Normative Ethics: On Algorithmic Bias and Disability, *First Monday*, (Jan. 2, 2023), 28:1.

[37] Matthew Harold Leiwant, "Locked Out: How Algorithmic Tenant Screening Exacerbates the Eviction Crises in the United States," 6 *Geo. L. Tech. Rev*. 276 (2022).

[38] Casey Ross & Bob Herman, "Denied by AI: How Medicare Advantage Plans Use Algorithms to Cut Off Care for Seniors in Need," *STAT News*, (Mar. 13, 2023)..

[39] Complaint of Disability Rights Washington, Self Advocates in Leadership, The Arc of the United States, and Ivanova Smith Against the Washington State Department of Health (WA DOH) and others, filed with HHS Office for Civil Rights, (Mar. 23, 2020), pp. 13-14.

[40] People with differing characteristics such as race and LGBTQ+ status are disproportionately associated with disability, but are generally recognized as situations linked to unequal health care access and unmet health needs. See, for example, Abigail Mulcahy, *et al*., "Gender Identity, Disability, and Unmet Healthcare Needs among Disabled People Living in the Community in the United States," *Int. J. Environ. Res. Public Health* February 23, 2022 (2022) 19:(5), 2588; National Congress of American Indians, Disabilities.

[41] The Center for Democracy and Technology and the eHealth Initiative published the *Proposed Consumer Privacy Framework for Health Data* in February 2021. The authors propose a voluntary self-regulatory program for non-HIPAA regulated companies that use, access, and disclose health data. Their framework has many detailed, thoughtful ideas, and they also stress that "this proposal is not designed to be a replacement for new and necessary comprehensive data privacy legislation."

[42] Deven McGraw & Kenneth D. Mandl, "Privacy Protections to Encourage Use of Health-Relevant Digital Data in a Learning Health System," *NPJ Digit. Med.* (Jan. 4, 2021), 4, 2.