American Business Law Journal

ACADEMY OF LEGAL STUDIES IN BUSINESS

American Business Law Journal
Volume 53, Issue 2, 251–314, Summer 2016

Healthy Predictions? Questions for Data Analytics in Health Care

*Janine S. Hiller**

Introduction

The Patient Protection and Affordable Care Act ("Affordable Care Act" or ACA),[1] health information technology (HIT) adoption, and increasing implementation of electronic medical records, are all propelling health care into the world of big data.[2] Big data, analytics, and predictive algorithms are poised to play a large part in the transformation of health-care delivery in the United States, determining who will benefit and, unfortunately, who may suffer from its insights. Health-care reform depends on cost savings derived from the application of sophisticated data analytics to the ever-expanding mass of data collected from and about individual patients. Health data analytics can lead to improved care, new scientific discoveries, and better medical treatment. Encouraging healthy behaviors, eliminating health disparities, and addressing the underlying determinants of health in society are important national goals.[3]

It is unclear, however, whether massive data collection about personal health and individual social status, both within the health-care system and outside of it, will serve the goal of addressing historical discrimination in health care, or whether data analytics will lead to the loss of

*R.E. Sorensen Professor in Finance, Professor of Business Law, Virginia Tech. The author thanks the Center for Business Intelligence Analytics, Pamplin College of Business, Virginia Tech, for its support of this research. The author thanks the participants at the 2014 Law and Predictive Analytics Conference at Virginia Tech for their helpful comments.

[1] Pub. L. No. 111-148, 124 Stat. 119 (2010).

[2] *See infra* text accompanying notes 91–92 for a description of big data.

[3] *See* U.S. Dep't Health & Human Servs., Healthy People 2020 Framework 1, https://www.healthypeople.gov/sites/default/files/HP2020Framework.pdf.

individual privacy, unequal treatment of individuals, and the perpetuation of health inequality. Data amassed from electronic health records (EHRs), private sector health website visits, personal health devices, mobile health applications, and social networks, are being linked together in a big data environment. Secondary use of health data by employers, insurers, marketers, and others heightens concerns. The collection and use of massive amounts of data about individuals, fed into a fragmented health analytics framework, may impose personal and societal costs if not carefully constructed. Furthermore, a predictive analytics[4] environment in health care may affect some groups differently than others, not decreasing health disparities but segmenting populations and resulting in differential care. As one author succinctly summarized, “An era of ‘big data’ promises exhilarating and frightening opportunities to cure and exploit human vulnerabilities.”[5] Health-care providers and policy makers should ask hard questions about how harms to personal privacy can be avoided, stigmas prevented, and threats of unbridled commercialization ameliorated.

This article proceeds in five parts. Part I provides a policy overview of health care and data. In order to examine the evolving, complex issues surrounding health data, analytics, and relationships to determinants of health, equality, and privacy, Part II broadly outlines use of health-care data from a policy perspective, including the ramifications of health-care reform for the collection of personal health data. Because the success of reform measures depends uniquely on the use of health data, Part III reviews a variety of methods for data collection from public and private sources, mobile devices, and social media, and then considers its use in health analytics. Part IV examines the laws and regulations that aim to both protect health data and patient privacy and to prohibit discrimination in health care.

Part V discusses the dynamic interplay between three aspects of today’s personal health data environment that create strong pressure on personal privacy and health priorities in ways that have not been critically acknowledged. First, policy goals, on the one hand, depend heavily on the use of big data to solve financial costs and population health problems and, on the other hand, promise to deliver equality in health care and to protect personal privacy, yet they fail to incorporate specific

[4] *See infra* text accompanying note 142 for a definition of predictive analytics.

[5] Frank Pasquale, *Grand Bargains for Big Data: The Emerging Law of Health Information*, 72 Md. L. Rev. 682, 684 (2013).

policies that will do both. Data take priority. Second, due to the emphasis on data driven health care, participation of commercial entities outside of the traditional health-care industry is mushrooming, as entities as diverse as data brokers and consulting firms collect and manage health data from within and outside the health system. Responding to the call for health analytics can both assist in promoting cures and deliver efficiencies, but the aggregation and manipulation of individual health data is occurring in ways that make it impossible for individuals to control its reach, and laws are inadequate to provide for robust privacy and antidiscrimination protection.

Lastly, the combination of policy that treats data as a solution and increasing data fusion threatens to turn health equality into data-driven discrimination. These thorny problems are not subject to easy solutions, but recognizing the problems can lead to intentional actions to avoid the most harmful consequences. Potential avenues for avoiding the unintended consequences of emphasizing data for health-care decisions, and for protecting individual privacy, are suggested.

I. Health Care and Data: Policy Overview

The U.S. health-care system is known as one of the most expensive and yet least efficient systems in the developed world, characterized before recent reform as a disparate have or have-not system without universal health care for all.[6] Decreasing the disparity in access to medical care between populations is probably the most commonly discussed aspect of health-care reform, and it is addressed significantly by extending the health-care mandate.[7] However, individual health in the United States also depends to a large degree on social constructs that lurk behind the health-care system, known as the social determinants of health; these

[6] *See* Lenny Bernstein, *Once Again, U.S. Has Most Expensive, Least Effective Health Care System in Survey*, Wash. Post: To Your Health (June 16, 2014), http://www.washingtonpost.com/news/to-your-health/wp/2014/06/16/once-again-u-s-has-most-expensive-least-effective-health-care-system-in-survey/ (presenting survey data before the health-care reforms).

[7] This aspect of the ACA, while critically important, is not the focus of this article. It has received considerable attention with regards to expanding access to underserved populations. *See, e.g.*, Rene Bowser, *The Affordable Care Act and Beyond: Opportunities for Advancing Health Equity and Social Justice*, 10 Hastings Race & Poverty L.J. 69, 79–95 (2013) (describing the opportunities and challenges for expanded care).

are the causes of the causes and they are related to race, ethnicity, income, education, and other social factors.[8] For example, a recent study by Johns Hopkins researchers found that contrary to previous assumptions, asthma affects both city and rural children at similar rates; the shared risk factors, however, are poverty, race, and ethnicity.[9] A brief background of the determinants of health and health-care disparities is necessary in order to establish its relationship to health-care reform and big data policies.

A. Determinants of Health and Health-Care Disparities

Determinants of health can be generally categorized as "genetics, behavior, social circumstances, environmental and physical influences, and medical care,"[10] although other categories may be used as well.[11] Thus, access to medical care is but one of the factors that can improve personal health status, and studies of the various determinants abound.[12] Recently, studies have focused on behavioral determinants that *undermine* health, such as smoking, obesity, and drug use.[13] But focusing only on individual behaviors misses the mark, as

> It is not an accident that people consume diets high in saturated fat and salt. It represents the nature of the food supply, culture, affordability, and availability, among other influences. *These are the causes of the causes*. . . . [For example,] in many rich countries now, there is a social gradient in smoking: the lower the socio-economic position, the higher the rate of smoking.[14]

[8] *See infra* Part II.A.

[9] *See Time to Rethink the Inner-City Asthma Epidemic?*, Johns Hopkins Children's Ctr. (Jan. 20, 2015), http://hopkinschildrens.org/Inner-City-Asthma-Epidemic/.

[10] Health Affairs, Health Policy Brief: The Relative Contribution of Multiple Determinants to Health Outcomes 2 (2014), http://www.rwjf.org/content/dam/farm/articles/articles/2014/rwjf415185.

[11] *See, e.g.*, Inst. of Med., Capturing Social and Behavioral Domains in Electronic Health Records: Phase 1, at 30–32 (2014) [hereinafter Social and Behavioral Domains] (describing different models, including the public health model, of determinants of health).

[12] *See* Health Affairs, *supra* note 10, at 4–5.

[13] *See id.*; *see also* Lindsay F. Wiley, *Shame, Blame, and the Emerging Law of Obesity Control*, 47 U.C. Davis L. Rev. 121, 131–41 (2013) (comparing obesity prevention to tobacco and HIV interventions).

[14] Michael Marmot, *Introduction*, *in* Social Determinants of Health 19, 21 (Michel Marmot & Richard G. Wilkinson eds., 2005) (emphasis added); *see also* Mary Anne Bobinski, *Health*

The World Health Organization, the Centers for Disease Control and Prevention, and the Robert Wood Johnson Foundation have all recognized the importance of understanding and addressing the social determinants of population health, in contrast to focusing purely on individual decisions and behaviors.[15] Social determinants of health contribute to health disparities, which can be defined as "a subset of health inequalities that are (i) based on factors such as socioeconomic status and racial/ethnic background and (ii) presumptively considered unjust."[16] Historically, discrimination and segregation in many aspects of U.S. health care were the norms.[17] Hospitals and nursing facilities were segregated, and patients were refused treatment if they were not Caucasian.[18] Differential care based on race and ethnicity persists today, and it is estimated that almost 85,000 deaths annually are related to racial bias in health care.[19] A 2003 report by the Institute of Medicine (IOM) to Congress reviewed over 100 studies that confirmed the disparity in care received due to racial status, taking into account other possible causes.[20] Physician bias, whether conscious or unconscious, results in different treatment for individuals based on their race, ethnicity, gender,

Disparities and the Law: Wrongs in Search of a Right, 29 Am. J.L. & Med. 363, 372–74 (2003) (discussing socioeconomic determinants of health).

[15] *See* Scott Burris, *From Health Care Law to the Social Determinants of Health: A Public Health Law Research Perspective*, 159 U. Pa. L. Rev. 1649, 1649–50 (2011). Whether it is within the jurisdiction of public health officials to address the broader societal determinants of health, or whether it is instead a series of political questions, is addressed in Micah L. Berman, *Defining the Field of Public Health Law*, 15 DePaul J. Health Care L. 45, 61–66 (2013).

[16] Emily Whelan Parento & Lawrence O. Gostin, *Better Health, but Less Justice: Widening Health Disparities After National Federation of Independent Business v. Sebelius*, 27 Notre Dame J.L. Ethics & Pub. Pol'y 481, 484 (2013) (also discussing differences in definitions).

[17] *See* Timothy Stoltzfus Jost, *Our Broken Health Care System and How to Fix It: An Essay on Health Law and Policy*, 41 Wake Forest L. Rev. 537, 611 (2006).

[18] *See* Sidney D. Watson, *Section 1557 of the Affordable Care Act: Civil Rights, Health Reform, Race, and Equity*, 55 How. L.J. 855, 860 (2012); Ruqaiijah Yearby, *When Is a Change Going to Come?: Separate and Unequal Treatment in Health Care Fifty Years After Title VI of the Civil Rights Act of 1964*, 67 SMU L. Rev. 287, 289 (2014).

[19] *See* Dayna Bowen Matthew, *Health Care, Title VI, and Racism's New Normal*, 6 Geo. J.L. & Mod. Critical Race Persp. 3, 4 (2014).

[20] Brian D. Smedley et al., Inst. of Med., Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care 38–77 (2003).

and disability status.[21] Some argue that socioeconomic status, including income and education, accounts for the majority of the disparity in health conditions in the United States.[22]

The lack of data to distinguish, compare, and determine causational relationships is one of the challenges for studying and tackling health disparities and determinants. In particular, problems in the past were both the lack of data to show that a disparity existed, and their longitudinal history.[23] The IOM summarized:

> Standardized data collection is also critically important in efforts to understand and eliminate racial and ethnic disparities in healthcare. Data on patient and provider race and ethnicity would allow researchers to better disentangle factors that are associated with healthcare disparities, help health plans to monitor performance, ensure accountability to enrolled members and payors, improve patient choice, allow for evaluation of intervention programs, and help identify discriminatory practices.[24]

The report also identified other limitations, such as non-standardized data and ethical concerns related to the collection of such data, including patient privacy concerns.[25]

The data element is a prickly component of population health and equality. In order to create the longitudinal and standardized data needed for addressing health disparities and determinants, it is necessary to collect individual-level health, social, and behavioral information. Health-care reform and the delivery of health care are related to this

[21] *See* Ruqaiijah Yearby, *Breaking the Cycle of "Unequal Treatment" with Health Care Reform: Acknowledging and Addressing the Continuation of Racial Bias*, 44 Conn. L. Rev. 1281, 1297–301 (2012) (describing medical and sociological studies that confirm unequal treatment).

[22] *See* Bobinski, *supra* note 14, at 372–73.

[23] *See* Stephen B. Thomas et al., *Less Talk More Action: Accelerating Innovative Strategies to Eliminate Racial and Ethnic Health Disparities*, 55 How. L.J. 705, 718 (2012); Watson, *supra* note 18, at 858.

[24] Inst. of Med., Unequal Treatment: What Healthcare Providers Need to Know About Racial and Ethnic Disparities in Healthcare 6 (2002), https://iom.nationalacademies.org/~/media/Files/Report%20Files/2003/Unequal-Treatment-Confronting-Racial-and-Ethnic-Disparities-in-Health-Care/Disparitieshcproviders8pgFINAL.pdf.

[25] *Id.* ("Unfortunately, standardized data on racial and ethnic differences in care are generally unavailable, and a number of ethical, logistical, and fiscal concerns present challenges to data collection and monitoring, including the need to protect patient privacy, the costs of data collection, and resistance from healthcare providers, institutions, plans and patients.").

data collection. A 2013 Special Communication in the *Journal of the American Medical Association* explained that "public health and social policy goals are applied to groups of people, not individuals."[26] An inherent tension arises when data are collected at the individual level in the clinical setting, and beyond. Even if it is de-identified and aggregated, in order to promote the personal healthy behaviors that are part of health-care reform, the data must be applied to individuals. The next section describes how health-care reform measures depend to a great degree on EHRs and how the use of health data is intended to promote the goals of an efficient, fair, and effective system.

B. Health-Care Reform and Data

The movement of paper health records to electronic health systems started the revolution in health-care data. In 2004, President Bush established a goal of fully implementing EHRs, and systems for sharing those records, by 2014.[27] By converting health information to electronic form, government medical reimbursements could move to more efficient electronic formats, fraud discovery and savings could be increased, and patient safety could be improved.[28] By 2013, great strides had been made toward achieving universal adoption of EHRs, as seventy-eight percent of physicians used some type of electronic health record system, and forty-eight percent of offices met the definition of a basic electronic system threshold.[29] Studies of sophisticated functionality of electronic

[26] *See* Hamilton Moses, III et al., *The Anatomy of Health Care in the United States*, 310 J. Am. Med. Ass'n 1947, 1961 (2013).

[27] *Promoting Innovation and Competitiveness: President Bush's Technology Agenda*, White House, http://georgewbush-whitehouse.archives.gov/infocus/technology/economic_policy200404/chap3.html (last visited Dec. 29, 2015) (setting ten-year goals).

[28] *See* Janine Hiller et al., *Privacy and Security in the Implementation of Health Information Technology (Electronic Health Records): U.S. and EU Compared*, 17 B.U. J. Sci. & Tech. L. 1, 1–3 (2011); Sharona Hoffman & Andy Podgurski, *Finding a Cure: The Case for Regulation and Oversight of Electronic Health Record Systems*, 22 Harv. J.L. & Tech 103, 112–19; Nir Menachemi & Taleah H. Collum, *Benefits and Drawbacks of Electronic Health Record Systems*, 4 Risk Mgmt. & Healthcare Pol'y 47, 47–48 (2011). In 2009, savings were estimated at seventy-seven billion dollars a year. *HIT or Miss*, Economist (Apr. 16, 2009), http://www.economist.com/node/13438006.

[29] Press Release, U.S. Dep't Health & Human Servs., More Physicians and Hospitals Are Using EHRs than Before (Aug. 7, 2014), http://www.hhs.gov/news/press/2014pres/08/20140807a.html.

records in hospitals reported fifty-nine percent adoption in the same year.[30] The U.S. Department of Health and Human Services (HHS) and the Office of the National Coordinator of Health Information Technology (ONC) continue to promote the adoption of EHRs and systems through a myriad of programs.[31] The federal government continues to invest heavily in EHRs to improve health care and reduce costs; for example, meaningful-use funding to incentivize use of EHRs will soon reach $47 billion.[32]

Although EHRs had their advent in the Bush administration, it was the passage of the ACA in 2010 that provided the impetus for a broader and more robust use of electronic health data for improvement of quality of care, clinical outcomes, and public health. A brief review of relevant parts of this law describes the shift in emphasis from the individual patient to population health outcomes and from medical interventions to behavioral changes—shifts that provide incentives to collect more individual health data from a myriad of sources and to use the data in more varied ways.

Two major goals of the ACA were to increase health-care coverage for Americans without insurance and to control and decrease spiraling health-care costs.[33] Related goals included focusing on (1) improving population health and (2) decreasing health disparities among

[30] *Id.*

[31] *See generally* https://www.healthit.gov/ (last visited Jan. 7, 2016).

> ONC is the principal federal entity charged with coordination of nationwide efforts to implement and use the most advanced health information technology and the electronic exchange of health information. The position of National Coordinator was created in 2004, through an Executive Order, and legislatively mandated in the Health Information Technology for Economic and Clinical Health Act (HITECH Act) of 2009.

About ONC, HEALTHIT.GOV, https://www.healthit.gov/newsroom/about-onc (last updated Aug. 11, 2014).

[32] *See* Deven McGraw & Alice Leiter, *A Policy and Technology Framework for Using Clinical Data to Improve Quality*, 12 HOUS. J. HEALTH L. & POL'Y 137, 138 (2012).

[33] Nat'l Fed'n of Indep. Bus. v. Sebelius, 132 S. Ct. 2566, 2580 (2012) ("The Act aims to increase the number of Americans covered by health insurance and decrease the cost of health care."); see also Mark T. Morrell & Alex T. Krouse, Accountability Partners: Legislated Collaboration for Health Reform, 11 Ind. Health L. Rev. 225, 235–36 (goals include improved health care for individuals, and populations, as well as reduced cost).

populations.[34] Data collection is key to accomplishing these goals in part because it provides the essential grist for creating analytical models that will predict which actions will improve health outcomes.[35] Consequently, savings from unnecessary and ineffective medical treatment can be used to expand services to those previously without access to the health-care system.[36] Furthermore, data collection about populations could gauge whether interventions to promote healthy behaviors were successful.[37]

Increasing population health was one reason the ACA created accountable care organizations (ACOs).[38] An ACO is an organization that is a patient's medical home, providing coordinated care among different medical professionals and practices.[39] The ACO agrees to accept a liquidated amount for aggregated patient care and is rewarded if overall costs are less than a predetermined amount.[40] Thus, the payment framework is based on healthy outcomes rather than numbers of medical interventions performed.[41] While some payments are tied to treatment for particular conditions, there are incentives for achieving broader-based, overall quality metrics and outcomes.[42] The payment policy also incentivizes the use of sophisticated data analytics to predict which patients will become ill; these models can be used either to

[34] *See* Eleanor D. Kinney, *The Affordable Care Act and the Medicare Program: The Engines of True Health Reform*, 13 Yale J. Health Pol'y, L. & Ethics 253, 292 (2013) (describing the Healthy People/Healthy Communities provision of the National Quality Strategy that resulted from the ACA).

[35] *See infra* Part III.B.

[36] *See* Nicolas P. Terry, *Protecting Patient Privacy in the Age of Big Data*, 81 UMKC L. Rev. 385, 414 (2012) (characterizing data analysis as being "hyped as the savior of health care").

[37] *See* Margaret B. Hoppin, Note, *Overly Intimate Surveillance: Why Emergent Public Health Surveillance Programs Deserve Strict Scrutiny Under the Fourteenth Amendment*, 87 N.Y.U. L. Rev. 1950, 1957–61 (2012) (describing a public health model for diabetes surveillance and follow-up, and privacy implications).

[38] *See* Jessica L. Mantel, *Accountable Care Organizations: Can We Have Our Cake and Eat It Too?*, 42 Seton Hall L. Rev. 1393, 1413–16 (2012).

[39] *Id*. at 1410.

[40] *See* Morrell & Krouse, *supra* note 33, at 244.

[41] *Id*. at 239–45.

[42] *See* Pasquale, *supra* note 5, at 735.

intervene for preventive purposes,[43] or, as an unintended consequence, to identify which patients health-care providers should avoid altogether.[44] In sum, the funding formula emphasizes rewards for promoting healthy behaviors rather than simply treatments for illness, preventative measures rather than medical interventions. Rewarding providers for maintaining healthy patients and emphasizing wellness also has external implications for patient and personal data collection, use, and sharing.[45] As financial rewards move away from numbers of patients treated towards numbers of patients who avoid treatment, doctors will find that "[*p*]*revention* requires tools that are often unfamiliar because educational, behavioral, and social interventions, not usually considered to be part of medicine, may be most effective for many diseases."[46] Thus, a consequence of expanding the realm of health care to include behavioral and social interventions for modifying patient behavior is the collection of detailed personal data in order to connect individual behavior to determinants of health.

While it is an essential component, universal access to health care will not by itself eliminate disparity of health-care treatment, which in the United States has a long history.[47] The ACA defines a health disparity in a population by reference to the Public Health Service Act, where it is determined that "there is a significant disparity in the overall rate of disease incidence, prevalence, morbidity, mortality, or survival rates in the population as compared to the health status of the general population"[48]

[43] *See* Sai T. Moturu et al., *Predictive Risk Modelling for Forecasting High-cost Patients: A Real-world Application Using Medicaid Data*, 3 INT. J. BIOMEDICAL ENGINEERING & TECH. 114, 115–16 (2010).

[44] *See* J. Frank Wharam & Jonathan P. Weiner, *The Promise and Peril of Healthcare Forecasting*, 18 AM. J. MANAGED CARE e82, e83 (2012) ("Some predictive model vendors openly acknowledge that their forecasting tools can be used to avoid high-risk patients or to identify those that will remain healthy.").

[45] *See* Mantel, *supra* note 38, at 1416–417 (noting that proponents argue that ACOs will have more resources for EHRs and will use the data to drive patient and treatment decisions). For a discussion of the broad extent of information collection and analytics, see *infra* Part III.

[46] *See* Moses et al., *supra* note 26, at 1947.

[47] *See supra* Part I.A.

[48] Affordable Care Act § 10334(c); 42 U.S.C. § 285t(d)(1) (2012).

To that end, the ACA includes several provisions that require that data be collected about a patient's race, ethnicity, sex, and language.[49] Population surveys, Medicaid, the Children's Health Insurance Program, and federally funded health programs must report data that include demographic information.[50] HHS implementation guidance for data collection sets a minimum level of information collection and standardized formats for race, ethnicity, sex, primary language, and disability status, stating that "[d]ata improvement efforts enhance the ability of the public health and healthcare systems to identify and track disparities in health and health care, understand their correlates and consequences, and facilitate greater accountability for reducing them."[51] The ACA creates a Community Preventive Services Task Force to study methods for reducing negative social determinants of health and to report on these remediation efforts every five years.[52]

Some medical professionals have proposed that social data also be collected and stored in the EHR so that it can be "used to affect medical risk and treatment decisions and to inform interventions to improve vulnerable patients' social circumstances."[53] The ethics of data collection and patient privacy were identified as requiring further attention, but a discussion of the details was absent.[54]

In summary, HIT has an expanded role to play in the future of health care: to enable the sharing of knowledge about individual treatments (in the EHR), to provide data for analytics, and to provide the fuel for predictions from which to create quality standards. The use of

[49] *See* Affordable Care Act. § 4302; 42 U.S.C. 300kk (2012).

[50] *See* Dennis P. Andrulis et al., Patient Protection and Affordable Care Act of 2010: Advancing Health Equity for Racially and Ethnically Diverse Populations 2 (2010), http://www.nashp.org/wp-content/uploads/sites/default/files/files/webinars/joint.center.ppaca_.health.equity.report.pdf.

[51] U.S. Dep't of Health and Human Servs., Implementation Guidance on Data Collection Standards for Race, Ethnicity, Sex, Primary Language, and Disability Status 1 (2011), https://aspe.hhs.gov/sites/default/files/pdf/76331/index.pdf.

[52] *See* Andrulis et al., *supra* note 50, at 13 (identifying transportation and environmental factors as examples of such determinants).

[53] *See* Laura Gottlieb et al., *Collecting and Applying Data on Social Determinants of Health in Health Care Settings*, 173 J. Am. Med. Assoc. 1017, 1018 (2013) (proposing also colocation of social and medical services such as legal aid).

[54] *Id*. at 1019.

HIT for electronic patient records provides a digital platform that enables the collection, aggregation, and ultimately the sharing of data and the formation of analytic applications and models. The end goal of the ACA and HIT is to repurpose health information from the individualized patient clinical record into a broader systemic health application in order to promote a "learning" health system.[55] In addition, comparative effectiveness research "require[s] individual- and population-level data that allow consideration of diagnostic, genetic, and ethnic differences; assessment of subpopulations; attention to current medical conditions and quality-of-life circumstances—and very large data sets."[56]

C. Big Data Policy and Health

In an increasingly data driven society, the benefits of collecting and using personally identifiable information for a social good, such as improvement of health care or increased security, are highly touted; yet the costs are often opaque. In May 2014, the White House released two reports in response to President Obama's request for a ninety-day study of the effect of big data on society.[57] Health-care data analytics were specifically referenced in these broader discussions.

1. "Seizing Opportunities, Preserving Values"

"Big Data: Seizing Opportunities, Preserving Values"[58] ("Big Data Report") was the product of a committee led by John Podesta, Counselor to the President, and included the Secretary of Commerce,

[55] *See* McGraw & Leiter, *supra* note 32, at 139–40 ("The vision is to create a 'learning' health care system that leverages clinical information in EHRs to improve the knowledge base about effective prevention and treatment strategies, and to disseminate that knowledge more quickly and efficiently to clinicians and patients to improve the quality and efficiency of health care.").

[56] Douglas Peddicord et al., *A Proposal to Protect Privacy of Health Information While Accelerating Comparative Effectiveness Research*, 29 Health Affairs 2082, 2084–85 (2010); *see also* Sharona Hoffman & Andy Podgurski, *The Use and Misuse of Biomedical Data: Is Bigger Really Better?*, 39 Am. J.L. & Med. 497, 499 (2013) (suggesting ACA comparative effectiveness research emphasis could lead to increased use of analytics).

[57] *See infra* Part II.C.1 & C.2.

[58] Exec. Office of the President, Big Data: Seizing Opportunities, Preserving Values (2014), https://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf [hereinafter Big Data Report].

Secretary of Energy, the Director of the Office of Science and Technology Policy, and the Director of the National Economic Council.[59] The committee focused primarily on the use, rather than collection, of big data, stating, "What really matters about big data is what it does," and consequently "whether our legal, ethical, and social norms are sufficient to protect privacy and other values in a big data world."[60] The Big Data Report notes an "asymmetry of power" between those who hold the data and those who supply it, and it notes the importance of understanding the many different contexts in which big data comes into play.[61] Three possible lenses for viewing the products of data and its analysis are identified: as private property, as a type of public good, or as an expression of individuality.[62] Situating data use within different contexts, the Big Data Report weaves normative, ethical, and legal considerations into the discussion, with the caveat that data technology cannot be viewed in isolation of its uses in society, and fundamental values should not be sacrificed to big data promises.[63]

The Big Data Report describes two institutional success stories of health analytics—a hospital program that analyzed data from minute-by-minute monitoring in a neonatal unit to predict future infections from a rise in temperature and heart rate, and a predictive analytics program used by the Centers for Medicare and Medicaid Services to identify fraud.[64] Another example was used to illustrate researchers' need for massive amounts of data in order to achieve results with health analytics. In this case, a substantial quantity of "particularly personal sensitive data like genetic data" was sought to obtain insight for purposes of treating schizophrenia, but privacy laws inhibited the collection of such data.[65] This example was used to highlight obstacles to

[59] *Id.* at iii.

[60] *Id.* at 3.

[61] *Id.*

[62] *Id.* at 3 n.9.

[63] *Id.* at 10.

[64] *Id.* at 6.

[65] *Id.* at 7.

researchers rather than to note intrusions on or harm to individual privacy and autonomy.

An entire section of the Big Data Report, entitled "Big Data and Health Care Delivery,"[66] strikes a positive stance toward the aggressive use of big data in the delivery of health care. Noting the emergence of EHRs, payment based on performance rather than service, in which positive outcomes are rewarded and "effective practices are identified from clinical data and then rapidly disseminated back to providers," the report concludes that big data will enable reimbursement based on "quality of patient outcomes."[67] Predictive medicine is described as the "ultimate application of big data in health,"[68] as the "powerful technology peers deeply into a person's health status and genetic information"[69] in order to forecast whether an individual will contract a disease or respond to a particular treatment. In order to develop this kind of personalized medicine, the report anticipates a change in approach to health privacy. Auspicious in its breadth, the report notes that because "lifestyle, genomic, medical, and financial data" are needed for the production of predictive health analytics, consequently "the distinction between personal data and health-care data has begun to blur."[70] The Big Data Report supports global access to medical records by researchers and those in the health system but notes that additional legislative protection may be necessary as private health information ultimately makes its way outside of traditional health entities.[71] The Big Data Report concludes with recommendations, one of which is to begin consultations about how the Health Insurance Portability and Accountability Act (HIPAA)[72] and other laws can be modified to allow for widely shared medical records in order to promote medical research and cost

[66] *Id.* at 22.

[67] *Id.* at 23.

[68] *Id.*

[69] *Id.*

[70] *Id.*

[71] *Id.* at 24.

[72] Pub. L. No. 104-191, 110 Stat. 1936 (1996).

savings in the health-care system.[73] A brief comment in the recommendations states that legal reforms regulating entities not presently covered by HIPAA should also be considered.[74]

The clear emphasis of the Big Data Report is on the inevitability and promotion of increased personal information collection as a precursor for sharing personal medical information, using big data and predictive analytics for medical research and treatment, and generating cost savings.

2. "A Technical Perspective"

The second report—"Big Data and Privacy: A Technological Perspective,"[75] ("PCAST Report")—written by the President's Council of Advisors on Science and Technology, likewise recognizes the potential of big data to deliver significant benefits.[76] It premises its recommendations on the assumption that technology does not stand in isolation of policy and cannot alone solve the big data problems.[77] The PCAST Report views technology benefits and harms as interconnected, through the lens of current values.[78] A health data example is used to illustrate the relationship between big data's potential for both benefits and harm, beginning with the acknowledgment that "powerful data analytics" can create inferences about individual health factors, thereby improving treatment.[79] In contrast, applying analytics (especially to nonclinical information) can infer information that previously would be considered to be private, and may be "offensive to widely shared values."[80] The report recognizes that big data analytics is not a perfect

[73] Big Data Report, *supra* note 58, at 62.

[74] *Id.*

[75] President's Council of Advisors on Sci. and Tech., Exec. Office of the President, Big Data and Privacy: A Technological Perspective (2014) [hereiafter PCAST Report], https://www.whitehouse.gov/sites/default/files/microsites/ostp/PCAST/pcast_big_data_and_privacy_-_may_2014.pdf.

[76] *Id.* at 11–14 (providing examples in general and for health care in particular).

[77] *Id.* at 48.

[78] *Id.* at 7.

[79] *Id.*

[80] *Id.* at 8.

solution or predictive vehicle; false positives and negatives, imprecise correlations, and the absence of causation can each mar the value of otherwise supportable statistical applications.[81]

The PCAST Report notes that individuals may not be harmed by imprecise analytics in certain circumstances, such as the "incorrect inference of a movie preference; *or the suggestion that a health issue be discussed with a physician*."[82] The report uses these two examples together, implying that neither will cause harm to an individual and conveying an implicit assumption that the overall positive impact of health analytics will outweigh the potential harm of incorrect or imprecise health analytics.

The PCAST Report also discusses anticipated uses of health care and analytics in the near future.[83] Personalized medicine based on genomics and the analysis of huge databases, and mobile device surveillance of health and health-related life activities, are two examples offered by the report.[84] The development of technical methods to protect privacy are specifically anticipated—allowing for the collection of individual identification but protecting access to health information and patient privacy.[85] On the other hand, the report identifies important questions about individualized, augmented, personal health care delivered via mobile device. Despite its potential to help users navigate health conditions, the third party use of such information, marketing, and lack of individual control over the highly sensitive information poses questions[86] that should be answered by policy and law.[87] Data and technical advances rated as probable, yet not predicted for the near future, describe a data driven, sensor filled world[88] in which individuals are intimately surveilled by devices in every room of their house and are "nudged" to

[81] *Id.*

[82] *Id.* (emphasis added).

[83] *Id.* at 13–14.

[84] *Id.* at 13.

[85] *Id.*

[86] *Id.* at 13–14.

[87] *Id.* at 47.

[88] *Id.* at 22.

share information in exchange for convenience and security.[89] It is not difficult to imagine this futuristic scenario applied to health and medical care, where surveillance prompts individuals in real time to make choices that promote healthy and cost effective lifestyles. The breadth of anticipated individual information collection, aggregation, and analytics that is identified in the PCAST Report has important implications for the privacy of personal health information. Privacy will be difficult to respect, as individual health information migrates out from the health system or is aggregated from unregulated sources. The report notes that requiring individual notice and consent for collection and use of data in the big data environment "is defeated by exactly the positive benefits that big data enables: new, non-obvious, unexpectedly powerful uses of data. It is simply too complicated for the individual to make fine-grained choices for every new situation."[90] This conclusion is particularly applicable to health care and the myriad of ways that have developed to collect and analyze sensitive, personal health and social information.

II. Data Analytics and Health Care

The burgeoning collection of patient data and adoption of EHRs are part of a larger environment of mushrooming data collection, storage, analytics, inference, and knowledge creation. This modern data is often identified as "big data." Mayer-Schönberger and Cukier identify three elements of the big data environment: access to huge amounts of data; use of messy rather than clean data; and the use of large, messy data sets to discover correlations and make predictions.[91] Definitions of big data and related concepts are not entirely uniform; however, these three elements provide a basis for understanding the uniqueness of the big data environment, where discovering correlation rather than

[89] *Id*. at 15–18.

[90] *Id*. at 38.

[91] *See* Viktor Mayer-Schönberger & Kenneth Cukier, Big Data: A Revolution That Will Transform How We Live, Work, and Think 19 (2013).

understanding causation is the knowledge goal.[92] The following sections describe the present and future of big data in the health-care field.

A. The Cascade of Health Data

As described in this part, structural changes to health-care provision in the United States are based in large part on the assumption that electronic health information can be combined and analyzed to provide insights to produce knowledge about better health outcomes, spur operational efficiencies, and predict which patients need care so as to prevent illness and blunt spiraling health costs. The implementation of EHRs may make possible the dismantling of legendary silos of health data held by providers.[93] Future adoption of Health Information Exchanges (HIEs) will encourage the sharing of health data across health systems.[94] As an example, a hospital traditionally had hundreds of different information systems that are not interoperable; it would not be unusual for the MRI to have its own proprietary information system that could not interact with the x-ray system that could not communicate directly with the doctor's system for patient care. It was very difficult for doctors to determine, for example, if ordering an MRI for a particular illness was effective, because the data's varying structure and

[92] *See* Simmi P. Singh & Tia Goss Sawhney, *Predictive Analytics and the New World of Retail Healthcare*, 27 Health Mgmt. Tech., Jan. 2006, at 46, 48 (noting that predictive analytics establishes a relationship, not causation).

[93] *See* Nicolas P. Terry, *Information Technology's Failure to Disrupt Health Care*, 13 Nev. L.J. 722, 746–48 (2013) (noting that disaggregation of health information from silos is necessary for success, but disruption of the existing models is difficult).

[94] *See* Ryan Abbott, *Big Data and Pharmacovigilance: Using Health Information Exchanges to Revolutionize Drug Safety*, 99 Iowa L. Rev. 225, 229–30 (2013) (stating that health information exchanges "will generate a never-before-seen amount of clinical data"); John W. Hill et al., A Proposed National Health Information Network Architecture and Complementary Federal Preemption of State Health Information Privacy Laws, 48 Am. Bus. L.J. 503 (2011) (arguing for a federal architecture to preempt state privacy law); Deth Sao et al., Interoperable Electronic Health Care Record: A Case for Adoption of a National Standard to Stem the Ongoing Health Care Crisis, 34 J. Legal Med. 55, 72–73 (2013) (describing ONC's goal of a National Health Information Network to share information, and the accompanying challenges). But see Nicolas P. Terry, Pit Crews with Computers: Can Health Information Technology Fix Fragmented Care?, 14 Hous. J. Health L. & Pol'y 129, 188 (2014) (asserting that it is difficult to determine whether information technology can help fix the health-care system or whether the health-care system must first be changed before taking advantage of the promises of health information technology).

location made it unobtainable and unusable. In the not too distant past, this health information would simply be discarded. EHRs can gather health information in one place, in the patient's record, thus creating a data environment that can be used for patient care and safety, research, and efficiency measures.[95] Data integration occurs when this information is fed into a "learning" system to better deliver health care; thus, the advent of big health data.[96]

Individual surveillance and resultant big data being fed into healthcare applications is not a big jump from the present day integration of data from multiple sources. EHRs contain an enormous amount of health data, and adoption by physicians and doctors continues to increase. In addition to the adoption of EHRs that fuels the growth of health data, additional clinical information from pharmacies, labs, health payers, and related health services feeds into the data stream.[97] Outside of the health professions, biometric monitoring devices (such as blood pressure or heart monitors), individually created information (for

[95] *See* Ranjit Janardhanan, *Uncle Sam Knows What's in Your Medicine Cabinet: The Security and Privacy Protection of Health Records Under the HITECH Act*, 30 J. Marshall J. Info. Tech. & Privacy L. 667, 702–03 (2014) (stating that a centralized information system is necessary for sharing information but may contribute to identity theft). For further analysis of legal issues related to health information technology and EHRs, see Leslie P. Francis, *When Patients Interact with EHRs: Problems of Privacy and Confidentiality,* 12 Hous. J. Health L. & Pol'y 171 (2012); Daniel J. Gilman & James C. Cooper, *There Is a Time to Keep Silent and a Time to Speak, the Hard Part Is Knowing Which Is Which: Striking the Balance Between Privacy Protection and the Flow of Health Care Information*, 16 Mich. Telecomm. & Tech. L. Rev. 279 (2010); Nicolas P. Terry & Leslie P. Francis, Ensuring the Privacy and Confidentiality of Electronic Health Records, 2007 U. Ill. L. Rev. 681 (2007).

[96] An Institute of Medicine report defined a learning system this way: "Advances in computing, information science, and connectivity can improve patient-clinician communication, point-of-care guidance, the capture of experience, population surveillance, planning and evaluation, and the generation of real-time knowledge—features of a continuously learning health care system." Inst. of Med., Best Care at Lower Cost: The Path to Continuously Learning Health Care in America 16 (Mark Smith et al., eds., 2012). *But see* Terry, *supra* note 94, at 146–58 (stating that significant barriers and market failures must be addressed in order for health information technology to reach its potential).

[97] *See* Allan F. Simpao et al., *A Review of Analytics and Clinical Informatics in Health Care*, 38 J. Med. Sys. 45, 46 (2014). One commentator defines four sources of health information: drugs and devices, clinical, financial and claims, and patient behavioral and sentiment based. *See* Terry, *supra* note 36, at 392–94. Of these sources, Terry considers the fourth to be potentially most harmful as it is outside a privacy-protecting framework and is subject to commercialization pressures. *Id.*

example, by Fitbit[98]), socioeconomic factors, environmental factors, and social media data can all be incorporated into health-related big data.[99] Data scientists consider Twitter posts, blogs, and status updates on Facebook within the realm of big data for health care.[100] This information is arguably relevant because it is estimated that medical intervention determines only ten to twenty percent of an individual's health status, with the remainder related to nonmedically determined environmental factors.[101] Yet, the intensive and longitudinal collection of patient data also seems particularly sensitive and intrusive.

Commercial entities are ready to assist the health community to take advantage of big data. IBM describes big data's application to health care this way:

> At the core of a data-driven healthcare organization is the ability to analyze a wide range of big data, from within and outside its four walls, to determine what is happening right now with regard to patient, staff and population profiles, as well as financial, clinical and operational processes. Big data comprises much larger volumes, wider varieties and greater velocities of data than most organizations have previously captured, stored and analyzed. It includes data from traditional sources such as electronic medical records (EMRs) and from nontraditional sources such as social media and public health records.[102]

Furthermore, in 2013, the McKinsey & Company report, "The 'Big Data' Revolution in Healthcare," identified the sources and applications of big data as clinical data, claims and cost data, pharmaceutical data, and "[p]atient behavior and sentiment data that describe patient

[98] Fitbit is one of many devices that track physical activity, calories burned, sleep habits, and other data. *See* Scott R. Peppet, *Unraveling Privacy: The Personal Prospectus and the Threat of a Full-Disclosure Future*, 105 Nw. U. L. Rev. 1153, 1154–55 (2011).

[99] *See* Bipartisan Pol'y Ctr., A Policy Forum on the Use of Big Data in Health Care 3 (2013), http://bipartisanpolicy.org/wp-content/uploads/sites/default/files/Use%20of%20-Big%20Data%20in%20Health%20Care.pdf.

[100] *See* Wullianallur Raghupathi & Viju Raghupathi, *Big Data Analytics in Healthcare: Promise and Potential*, 2 Health Info. Sci. & Sys. 3, 3 (2014) (discussing the varied sources of health information, both inside and outside the health care system).

[101] *See* Health Affairs, *supra* note 10, at 4–5 (summary of studies of the determinants of health).

[102] IBM, Data-Driven Healthcare Organizations Use Big Data Analytics for Big Gains 2 (2013), http://www-03.ibm.com/industries/ca/en/healthcare/documents/Data_driven_healthcare_organizations_use_big_data_analytics_for_big_gains.pdf.

activities and preferences, both inside and outside the healthcare context. . . ."[103] This last source of external health data is significantly different from the traditional sources of health information. The McKinsey report explained that "payors can learn about patients' finances, buying preferences, and other characteristics through companies that aggregate and sell consumer information, such as Acxiom and Accurint" [data brokers].[104] Acxiom is a data broker that has collected over 3000 data points for nearly every consumer in the United States and tracks information for approximately 700 million consumers globally.[105]

Data brokers, companies that collect broad swaths of personally identifiable information about nearly every individual, aggregated from different sources, create unique categories of persons that are available for purchase.[106] The World Privacy Forum discovered that lists of rape victims, AIDS sufferers, and individuals with dementia, were all available for sale.[107] A year later, a credit card website stated that "everyone from the health insurer Blue Cross Blue Shield to the major analytics company SAS to the data giant FICO is experimenting with or at least talking about using consumer data for health-care purposes,"[108] and noted that credit card purchase information is being used as part of a scoring

[103] Peter Groves et al., The 'Big Data' Revolution in Healthcare: Accelerating Value and Innovation 3 (2013), http://www.mckinsey.com/~/media/mckinsey/dotcom/client_service/Healthcare%20Systems%20and%20Services/PDFs/The_big_data_revolution_in_healthcare.ashx.

[104] *Id.*

[105] *See* Fed. Trade Comm'n, Data Brokers: A Call for Transparency and Accountability 8 (2014), https://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf [hereinafter FTC Data Broker Report].

[106] *See* Shannon Pettypiece & Jordan Robertson, *Sick Elderly for Sale by Data Miners for 15 Cents a Name*, BloombergBusiness (Sept. 11, 2014, 3:07 PM), http://www.bloomberg.com/news/articles/2014-09-11/how-big-data-peers-inside-your-medicine-chest.

[107] Melanie Hicken, *Data Brokers Selling Lists of Rape Victims, AIDS Patients*, CNN (Dec. 19, 2013, 12:38 PM), http://money.cnn.com/2013/12/18/pf/data-broker-lists/.

[108] Kelly Dilworth, *Health Care Companies Turn to 'Big Data'*, CreditCards (Aug. 14, 2014), http://www.creditcards.com/credit-card-news/health-care-companies-turn-to-big-data-1282.php.

system that can be applied to individual patients and predict their chance of becoming chronically ill.[109] A physician working on this project stated that "[w]hen we leverage the data, when we start to think about populations as a whole, we can really have huge impacts on patients' care."[110] The benefits and risks of applying a predictive system to treat individual patients, and the potential conflict between population and individual outcomes, are addressed below in Part IV.B.

The amount and variety of personal information included in health data streams is explained by the integration of data from data brokers and the increasing use of data produced by social media. "The totality of data related to patient healthcare and well-being" includes traditional clinical and test results, as well as individual patient "social media posts, including Twitter feeds (so-called tweets), blogs, status updates on Facebook and other platforms, and web pages."[111] Websites or Facebook pages devoted to patients with specific health problems, and their friends and family, are maintained by for-profit organizations that encourage individuals to share personal health successes and failures.[112]

One example is PatientsLikeMe, which ran a promotion of "24 Days of Giving" asking patients to "simply share their health data for good." [113] By creating a personal profile, tracking their health symptoms, treatments, quality of life, and making the information public, it proposed that "patients not only help themselves, but help others who can learn from their experiences, and advance research."[114] The site is integrated with Facebook and Twitter, thereby expanding the distance personal information will travel and increasing the probability that the information will be aggregated. A person who responds to the plea to contribute to the social

[109] *Id.*

[110] *Id.*

[111] *See* Raghupathi & Raghupathi, *supra* note 100, at 1 (footnotes omitted).

[112] For further discussion of interactive health websites and the sharing of patient information, see Deborah Lupton, *The Commodification of Patient Opinion: The Digital Patient Experience Economy in the Age of Big Data*, 36 Soc. of Health & Illness 856 (2014).

[113] Press Release, PatientsLikeMe, PatientsLikeMe Launches "24 Days of Giving" to Encourage Sharing of Health Data for Good (Nov. 17, 2014), http://news.patientslikeme.com/press-release/patientslikeme-launches-24-days-giving-encourage-sharing-health-data-good.

[114] *Id.*

good will likely be contributing information to a far wider circle than perhaps he or she anticipated, and benefiting purely commercial organizations as well as medical researchers. In addition to the specific promotion, individuals earn stars for sharing pieces of personal health information—three stars earn the person a t-shirt.[115] PatientsLikeMe's "Frequently Asked Questions" explains that it "[T]ake[s] the information patients like you share about your experience with the disease and sell[s] it to our partners"[116] in order to support its for-profit goal. A person is advised to "expect that every piece of information you submit (even if it is not currently displayed) may be shared with our partners . . . ."[117] PatientsLikeMe and its founders have earned awards for contributions to health improvement and patient engagement. Its core values include transparency and openness.[118] However, the potential harms that sharing health data and personal information could induce is also recognized at the site:

> It is possible that a Member could be identified using information shared on PatientsLikeMe (and/or in conjunction with other data sources). A Member could be discriminated against or experience repercussions as a result of the information shared. For example, it is possible that employers, insurance companies, or others may discriminate based on health information.[119]

Not surprisingly, neither the data use nor the patient warning information is found explicitly at the main page or at the promotional page that advocates for the social good of sharing. Instead, the foregoing is

[115] *What Do the Stars Mean? How Do I Get Three Stars (and a Free T-Shirt)?*, PatientsLikeMe, https://support.patientslikeme.com/hc/en-us/articles/201186684-What-do-the-stars-mean-How-do-I-get-three-stars-and-a-free-T-shirt- (last visited Dec. 29, 2015).

[116] *How Does PatientsLikeMe Make Money?*, PatientsLikeMe, https://support.patientslikeme.com/hc/en-us/articles/201245750-How-does-PatientsLikeMe-make-money- (last visited Dec. 29, 2015).

[117] *Id.*; *see also* Frank Pasquale & Tara Adams Ragone, *Protecting Health Privacy in an Era of Big Data Processing and Cloud Computing*, 17 Stan. Tech. L. Rev. 595, 633–34 (2014) (describing Nielsen copying the entirety of PatientsLikeMe online forums for data analytics).

[118] *What Are the Company's Core Values?*, PatientsLikeMe, https://support.patientslikeme.com/hc/en-us/articles/201245710-What-are-the-company-s-core-values- (last visited Dec. 29, 2015).

[119] *Privacy Policy*, PatientsLikeMe, http://www.patientslikeme.com/about/privacy (last visited Dec. 29, 2015).

provided as a link at the registration page where one must check "I agree" in order to enroll.[120]

1. Patient-Generated Data

As the previous discussion indicates, patient data in the era of "health 2.0" includes both a wide array of information from diverse sources and data maintained over time and connected with individual medical records. Patient-generated data are defined as "health-related data—including health history, symptoms, biometric data, treatment history, lifestyle choices, and other information—created, recorded, gathered, or inferred by or from patients or their designees (i.e., care partners or those who assist them) to help address a health concern."[121]

The ONC is considering including patient-generated health data in stage three of its meaningful use program for the promotion of HIT.[122] ONC seeks to involve patients in their care through education, self-monitoring, and self-management;[123] requiring health-care providers and electronic records to interact with social media and health information generated from patient self-monitoring would take this goal to another level. Privacy and security of patient information as well as limited access to the data are concerns, particularly as it relates to trust within the health system.[124] Data integrity is also important and is

[120] See the registration page found at https://www.patientslikeme.com/join. When this page was last visited on December 29, 2015, a user's personal endorsement for the site was found on the right hand side. The links for the privacy policy and the terms of use are also found on the main page of the website, but are in smaller type, and consist of blue links on a blue background.

[121] Mary Jo Deering, Issue Brief: Patient-Generated Health Data and Health IT 3 (2013), https://www.healthit.gov/sites/default/files/pghd_brief_final122013.pdf; *see also* Jodi Daniel et al., Office of Nat'l Coordinator for Health Info. Tech., Issue Brief: Using Health IT to Put the Person at the Center of Their Health & Care by 2020, at 4 (2014), http://www.healthit.gov/sites/default/files/person_at_thecenterissuebrief.pdf (listing as one 2020 goal of seamless interaction with the health-care system is to "[s]often or erase the boundaries between what occurs inside and outside of the health care system by promoting increased information flow"); Michael Shapiro et al., Patient-Generated Health Data 2 (2012), http://www.rti.org/pubs/patientgeneratedhealthdata.pdf (report prepared for the Office of Policy and Planning, Office of the National Coordinator for Health Technology).

[122] *See* Deering, *supra* note 121, at 6.

[123] *Id*. at 9.

[124] *Id*. at 10.

related to source identification and accuracy.[125] ONC notes further that there "could be a need to address patient authorization for secondary sharing . . . if the patient prefers that the data not be shared with other providers or for other purposes."[126]

On another front, in May 2014 the Federal Trade Commission (FTC) hosted a workshop on "Consumer Generated and Controlled Health Data."[127] The seminar participants provided information about how health information moves from patient to doctor and then is channeled to a considerable number of entities outside of regulatory oversight.[128] Thirty-three states share or sell health databases about hospital stays, and commercial entities are top purchasers in approximately a dozen states.[129] In addition, health information and data was discussed in the FTC's 2015 report, "Internet of Things: Privacy and Security in a Connected World," showing the inherent connection between patient generated information and health data collected through mobile applications.[130]

2. Mobile Health Data

Seventy percent of Americans keep tabs on at least one indicator of health—many using personal monitors, which increasingly track

[125] *Id.*

[126] *Id.*

[127] *Spring Privacy Series: Consumer Generated and Controlled Health Data*, FED. TRADE COMM'N (May 7, 2014), https://www.ftc.gov/news-events/events-calendar/2014/05/spring-privacy-series-consumer-generated-controlled-health-data.

[128] *See* FED. TRADE COMM'N, TRANSCRIPT, SPRING PRIVACY SERIES: CONSUMER GENERATED AND CONTROLLED HEALTH DATA 12–21 (2014), https://www.ftc.gov/system/files/documents/videos/spring-privacy-series-consumer-generated-controlled-health-data/ftc_spring_privacy_series_-_consumer_generated_and_controlled_health_data_-_transcript.pdf.

[129] *See* LATANYA SWEENEY, SPRING PRIVACY SERIES: CONSUMER GENERATED AND CONTROLLED HEALTH DATA, http://patientprivacyrights.org/wp-content/uploads/2014/06/consumer-health-data-webcast-slides.pdf (last visited Dec. 29, 2015).

[130] FED. TRADE COMM'N, INTERNET OF THINGS: PRIVACY AND SECURITY IN A CONNECTED WORLD 7–8, 14–16 (2015), https://www.ftc.gov/system/files/documents/reports/federal-trade-commission-staff-report-november-2013-workshop-entitled-internet-things-privacy/150127iotrpt.pdf [hereinafter FTC IoT] (discussing benefits and risks of consumer generated health information).

multiple aspects of individual health, such as sleep patterns and weight.[131] A recent example highlights the increasingly available mobile applications that make up what is called the mHealth environment. Moves is a phone app that incorporates GPS and motion sensors to determine where a person is, how far they have traveled, how many calories they have burned, and a map of the user's locations.[132] When Facebook acquired Moves, promises were made not to integrate the app with the popular social media site, but Facebook still had access to users' locations, and the Moves privacy policy was changed in order to allow sharing with third parties.[133] Privacy advocates predict that mobile health app companies will monetize the data that they collect and that "[i]nformation about consumers [sic] most intimate health conditions is going to be sold to the highest bidder."[134] This intimate information would likely end up in the hands of data brokers, and personal information would be packaged and sold in one of the hundreds of health categories that already exist, such as diabetics, sufferers of depression, "Suffering Seniors," and "Aching and Ailing."[135] In the first part of 2014, numbers of available health apps were estimated to top 100,000, and revenues for the apps and the services that go with them were projected to reach $26 billion by 2017.[136] A recent market study predicted the following:

> [N]ot only will new players such as sensor vendors or mHealth data aggregators enter the healthcare market, but . . . they will become the dominant

[131] *See* Moses et al., *supra* note 26, at 1959.

[132] *See* Lance Whitney, *Facebook Acquires Health and Fitness Tracking App Moves*, C|NET (Apr. 24, 2014, 8:09 AM), http://www.cnet.com/news/facebook-acquires-health-fitness-tracking-app-moves/.

[133] *See* Andrea Peterson, *Privacy Advocates Warn of 'Nightmare' Scenario as Tech Giants Consider Fitness Tracking*, WASH. POST. (May 19, 2014), http://www.washingtonpost.com/blogs/the-switch/wp/2014/05/19/privacy-advocates-warn-of-nightmare-scenario-as-tech-giants-consider-fitness-tracking/.

[134] *Id.*

[135] *See* Pettypiece & Robertson, *supra* note 106 (noting that names sell for fifteen cents apiece).

[136] *See* John N. Frank, *Mobile Health App Revenue to Grow Tenfold by 2017, Study Predicts*, MODERN HEALTHCARE (May 22, 2014), http://www.modernhealthcare.com/article/20140522/blog/305229997.

> participants. Traditional healthcare players need to understand what impact e.g. health API and data aggregators will have on their business models. A clear understanding of the growing connected mHealth app market is indispensable for those traditional healthcare players who do not wish to be left out from the new ecosystem.[137]

There are important potential benefits of mHealth data when it is shared with health-care providers. Management of diabetic treatment and high blood pressure by integration of the data from patient medical devices with health-care professionals could lead to better care,[138] for example. Patient management that is facilitated by the data from health-related apps is also believed to be capable of extending home care over institutional health care for patients.[139] On the other hand, a study of mobile health applications found that forty percent of commercial apps did not have a privacy policy; of those that did have a privacy policy, thirty percent violated the policy by sharing information with an undisclosed third party, and only ten percent encrypted all of the health information during mobile transmission.[140] The incredible potential benefits of some classes of mobile health data can be compared with the potential risks of classes of unregulated third-party mobile health apps.[141]

B. Predictive Health Analytics

A general definition of predictive analytics is that it is the application of a mathematical algorithm to large data sets resulting in a prediction of a

[137] RESEARCH2GUIDANCE, MHEALTH APP DEVELOPER ECONOMICS 2014, at 32 (2014), http://www.research2guidance.com/r2g/research2guidance-mHealth-App-Developer-Economics-2014.pdf.

[138] *See* FTC IOT, *supra* note 130, at 7–8.

[139] *Id*. at 7.

[140] *See* LINDA ACKERMAN, PRIVACY RIGHTS CLEARINGHOUSE, MOBILE HEALTH AND FITNESS APPLICATIONS AND INFORMATION PRIVACY 5 (2013), https://www.privacyrights.org/mobile-medical-apps-privacy-consumer-report.pdf (a report to the California Consumer Protection Foundation).

[141] *See generally* Nathan Cortez, *The Mobile Health Revolution?*, 47 U.C. DAVIS L. REV. 1173 (2014); Anne Marie Helm & Daniel Georgatos, *Privacy and MHealth: How Mobile Health "Apps" Fit into a Privacy Framework Not Limited to HIPAA*, 64 SYRACUSE L. REV. 131 (2014); Daniel F. Schulke, *The Regulatory Arms Race: Mobile-Health Applications and Agency Posturing*, 93 B.U. L. REV. 1699 (2013).

future occurrence.[142] Predictive analytics for health care can be used on a number of different levels, including health research, fraud detection, and clinical patient care. IBM forecasts that "analytics in healthcare is moving toward a model that will eventually incorporate predictive analytics and enable organizations to 'see the future,' create more personalized healthcare, allow dynamic fraud detection and predict patient behavior."[143] A survey of leading health organizations, however, indicated that the greatest use of data analytics was anticipated to be for sales and marketing.[144] Nevertheless, cost containment and efficient delivery of high-quality health care are common goals of predictive analytics.[145]

Health care will be ineffective and costs will rise, for example, if a patient will not take the prescribed medicine. Therefore, predicting patient behavior becomes a key aspect to be investigated by the use of analytics.[146] The application of health information analytics is expected to improve patient outcomes by predicting which individual patients would benefit from either medical intervention or lifestyle changes.[147] Using large amounts of data from clinical and nonclinical sources, analytic models can also predict who will become a high-risk patient, which is of particular interest to health-care providers, employers, and insurers.

[142] *See* I. Glenn Cohen et al., *The Legal and Ethical Concerns That Arise from Using Complex Predictive Analytics in Health Care,* 33 Health Affairs 1139 (2014).

[143] James W. Cortada et al., The Value of Analytics in Healthcare: From Insights to Outcomes 3 (2012), https://www.ibm.com/smarterplanet/global/files/the_value_of_analytics_in_healthcare.pdf.

[144] *Id*. at 4.

[145] *See infra* Part III.B.

[146] *See generally* Paul Bradley, *Predictive Analytics Can Support the ACO Model*, 66 Healthcare Fin. Mgmt. 102 (2012); Nitesh V. Chawla & Darcy A. Davis, *Bringing Big Data to Personalized Healthcare: A Patient-Centered Framework*, 28 J. Gen. Internal Med. S660, S661 (2013) ("[T]he problem of patient-centered and personalized disease risk profile is analogous to recommendation systems used for movies or books."); Peter Edelstein, *Emerging Directions in Analytics*, 34 Health Mgmt. Tech. 16, 17 (2013) (stating that predictive analytics can help "deliver the right message to the right patient at the right time using the right medium or technology").

[147] *See* Raghupathi & Raghupathi, *supra* note 100, at 2.

One recent example of health data and analytics is the collaboration between IBM, Epic, and Carilion Clinic health systems. Generally, Epic provides EHRs for Carilion, and IBM provides the data warehouse.[148] Carilion, an ACO, engaged the two businesses to use three years of data on 350,000 patients, including over twenty million clinical notations, to create a model to predict patients who were a future risk for heart failure.[149] The process identified 8500 at-risk patients; consequentially, Carilion will implement additional health interventions and track success rates of those interventions.[150] To create a risk factor for each individual patient, IBM took advantage of clinical notes relating to *nonclinical* information that might otherwise have been overlooked, including "a patient's social history, depression, and living arrangements."[151] A description of a general health-care analytics product from IBM describes the source of data for its predictive models as thousands of factors that produce "similarities and differences" between patients in order to suggest a personalized health approach to a provider in real time at the patient care decision point.[152] Clinical, demographic, and environmental factors are used in the comparisons, including individual "social, educational, and financial factors."[153]

In another example, Carolinas HealthCare System integrated data obtained from a person's general purchases from store loyalty cards with data from a data broker with health-care information in order to create a patient "risk score" that will be used to identify patients that

[148] *See* Zina Moukheiber, *IBM and Epic Apply Predictive Analytics to Electronic Health Records*, FORBES (Feb. 19, 2014, 1:59 PM), http://www.forbes.com/sites/zinamoukheiber/2014/02/19/ibm-and-epic-apply-predictive-analytics-to-electronic-health-records/.

[149] *Id.*

[150] *Id.*

[151] *Id.*

[152] *See* IBM SOFTWARE, A BETTER WAY TO DELIVER HEALTH AND SOCIAL CARE 4 (2013), http://www-01.ibm.com/common/ssi/cgi-bin/ssialias?infotype=PM&subtype=SP&htmlfid=ZZS03187USEN (the product is named "Patient Similarity Analytics"). Walgreens is using a similar point-of-care patient treatment tool created from over 8 billion medical data points. *Walgreens Brings Big Data Analytics to Healthcare Clinics Through Expanded Relationship with Inovalon*, BUSINESS WIRE (Jan. 30, 2014, 10:00 AM), http://www.businesswire.com/news/home/20140130005739/en/Walgreens-Brings-Big-Data-Analytics-Healthcare-Clinics#.VNkVw8ZY3i8.

[153] *See* IBM SOFTWARE, *supra* note 152, at 4.

should be contacted—*before* they suffer health problems.[154] Predictive models have discovered some interesting correlations from nonhealth information. For example, one hospital conglomerate in Pennsylvania found that Internet shoppers were more likely to use emergency services than non-Internet shoppers.[155]

In addition to the goal of identifying and reaching at-risk patients, there are other reasons to use predictive analytics. Some health-care providers seek to reach out to high income and "well-insured" patients in order to induce further use of the health-care system[156] and higher profits, or to increase the use of "revenue-generating areas" such as operating rooms.[157] Another profit related use of predictive analytics is its application to monitor health-care provider charges, so that every recoverable health cost item is discovered and billed.[158]

Furthermore, predictive health analytics can be used for population health decisions. For example, in order to reach broad goals for a certain population, "A [predictive] model may recommend withholding a potentially beneficial intervention from some patients with a given condition because there is a significantly lower probability that they will benefit, while offering the intervention to others who are more likely to benefit."[159] The allocation of health benefits based on a mathematical algorithm can contain data from a wide swath of sources, both within and without the health system, raising questions about equity, predisposed stereotypes, and discrimination based on factors such as race, sex,

[154] *See* Shannon Pettypiece & Jordan Robertson, *Doctors Soon See Donuts-to-Cigarette Charges for Health*, BloombergBusiness (June 26, 2014, 11:35 AM), http://www.bloomberg.com/news/articles/2014-06-26/hospitals-soon-see-donuts-to-cigarette-charges-for-health.

[155] *See* Natasha Singer, *When a Health Plan Knows How You Shop*, N.Y. Times (June 28, 2014), http://www.nytimes.com/2014/06/29/technology/when-a-health-plan-knows-how-you-shop.html?_r=0.

[156] *Id.*

[157] Health Fidelity identifies "enterprise performance"—i.e., maximizing resources—as one of the three most common uses of health analytics. *See* Health Fidelity, Unlocking the Value of Healthcare's Big Data with Predictive Analytics 3 (2013), http://docplayer.net/985071-Unlocking-the-value-of-healthcare-s-big-data-with-predictive-analytics.html.

[158] *See* Paul Bradley & Jeff Kaplan, *Turning Hospital Data Into Dollars*, 64 Healthcare Fin. Mgmt. 64, 64 (2010); Pieter Schouten, *Big Data in Health Care*, 67 Healthcare Fin. Mgmt. 40, 40 (2013).

[159] *See* Cohen et al., *supra* note 142, at 1139–40.

and financial status.[160] The aggregation of personal data with health data is arguably more intrusive, and as surveillance of individuals becomes part of the big data health picture, it raises further questions about whether data and predictions could be used outside the health-care domain to discriminate against individuals in areas such as employment, insurance, and education. The laws and regulations that might intercede in these results are addressed in the next part.

III. Protection of Individuals and Their Health Information

When President Bush announced the goal of implementing EHRs for the majority of U.S. citizens by the year 2014, individual concerns about the privacy and security of EHRs were viewed as an impediment to adoption, and subsequent surveys supported this perception.[161] But in addition to privacy questions, individuals expressed significant concern about medical identity theft, use of health information for marketing purposes, employers' access to health records, and insurers' access to health information.[162] Furthermore, as analytics are used to assess improvement of healthy outcomes for populations, questions of different treatment of citizens will arise. Both privacy and nondiscrimination laws become particularly relevant.

A. Privacy of Personal Health Information: HIPAA and HITECH

Although various state laws protect the privacy of health information,[163] the following discussion focuses on federal law that applies to health information in the hands of health-care providers—HIPAA, as amended in 2009 by the Health Information Technology for Economic and Clinical Health Act (HITECH).[164] A brief history of the relevant provisions

[160] *See infra* Part IV.B.

[161] *See* Hiller et al., *supra* note 28, at 6–7.

[162] *Id.*

[163] *See* Hill et al., *supra* note 94, at 521–31 (reviewing state laws applicable to health information privacy).

[164] Pub. L. No. 111-5, 123 Stat 115 (2009). HHS adopted regulations known as the Privacy and Security Rules. To simplify the discussion, these are referred to in the text under the

of these statutes and their implementation shows an attempt to address the expanding number of entities that handle health information and to bring them under a regulatory umbrella. Of particular importance to the cascading of health data is the delineation of protected information and the kinds of entities that are subject, or not subject, to the law.

HIPAA applies to protected health information (PHI), which is individually identifiable health data. This includes common information such as one's address, social security number, and the like, but could be any information that serves to reasonably identify a person.[165] It follows that data that does not include identifying information is not covered by HIPAA limitations.[166] Furthermore, if information is stripped of specified information, then it is de-identified, and not considered to be PHI.[167] De-identification standards are met when there is no "reasonable basis" on which to identify a person from the data.[168] This occurs when either a qualified statistician formally determines it to be true, or when certain information is removed about the person, relatives, members of the household, or employers.[169] The covered entity may not have knowledge of a means by which a person may be identified.[170]

statutory authorization, HIPAA or HITECH. Most recent modifications to the regulations occurred in 2013 and are commonly referred to as the Omnibus Privacy Rule. 45 C.F.R. §§ 160, 162, 164 (2015).

[165] 45 C.F.R. § 160.103 (2015).

[166] 45 C.F.R. § 164.514(a) (2015) provides, "Health information that does not identify an individual and with respect to which there is no reasonable basis to believe that the information can be used to identify an individual is not individually identifiable health information."

[167] 45 C.F.R. § 164.514(b). The ONC Guide to Privacy and Security of Electronic Health Information, version 2.0, released in April 2015 notes that "[i]f data is de-identified in the manner prescribed by HIPAA, it is not PHI. Increasingly researchers are seeking and using de-identified clinical data for health system improvement activities." Office of the Nat'l Coordinator for Health Info. Tech., Guide to Privacy and Security of Electronic Health Information 20, http://www.healthit.gov/sites/default/files/pdf/privacy/privacy-and-security-guide.pdf.

[168] 45 C.F.R. § 164.514(a).

[169] 45 C.F.R. § 164.514(b)(2).

[170] 45 C.F.R. § 164.514(b)(2)(ii).

Originally, only PHI held by "covered entities" was protected. This encompassed health plans, health-care clearinghouses, and health-care providers who transmit health information electronically in certain health information transactions.[171] This definition proved to be too narrow in light of numerous third party entities in possession of personal health information for operational and processing reasons. Subsequently, covered entities were required to ensure, by explicit agreement, that business associates used the information in the manner in which it was intended, that the data were secure, and that the business associate would cooperate to protect such data.[172] HITECH extended the reach even further to *directly* cover "business associates," defined as entities in possession of personal health information as a result of using PHI in the performance of their duties to covered entities. This includes personal health records and electronic data for PHI.[173]

Access to or sharing of PHI is based on the intended use of the data. A covered entity must obtain consent to share PHI if the entity sells the information for either direct or indirect compensation, or marketing.[174] Beyond these areas, a patient's right to restrict sharing of her data is quite limited. Without the patient's consent, a covered entity may only share the minimum necessary health information for purposes of treatment, payment, or health-care operations.[175] Incidental uses and situations that imply consent will allow the covered entity to disclose PHI. Furthermore, some scholars advocate for a change in the law in order to expand the right to use PHI for research, usually combined with an obligation to secure the information from misuse.[176] In contrast, it has been argued that a more exact list of information considered the

[171] *See* Hiller et al., *supra* note 28, at 11–12.

[172] *See* Janardhanan, *supra* note 95, at 697–99.

[173] *Id.* at 697. For additional discussion of which entities will fall under the business associate category, see Pasquale & Ragone, *supra* note 117, at 609–15.

[174] HITECH, § 13405; 42 U.S.C. § 17935 (2012).

[175] 45 C.F.R. § 164.502 (2015).

[176] *See* Sharona Hoffman & Andy Podgurski, *Balancing Privacy, Autonomy, and Scientific Needs in Electronic Health Records Research*, 65 SMU L. Rev. 85, 124–25 (2012) (asserting that requiring patients to share their EHRs is "ethically sound"); McGraw & Leiter, *supra* note 32, at 156–66 (describing a framework for secondary use of clinical data to be expanded under fair information practice principles); Suzanne M. Rivera, *Privacy vs. Progress:*

minimum necessary under specific medical circumstances is needed in order to limit sharing and strengthen protection of health information.[177]

Specific exceptions apply when the covered entity is *not* required to obtain patient consent.[178] For purposes of this discussion, the most relevant circumstance is when PHI is transformed into limited data sets. A limited data set is created when certain identifying information is removed. It may then be shared for "research, public health, or health care operations" if an agreement to safeguard such information is entered into with the third party to whom the limited data set is transferred.[179] A limited data set is different from de-identified data because it can include information that is useful for public health purposes, including birth date, treatment date, and more specific geographical identifiers.[180] Especially for limited data sets, which contain more specific information, it has been argued that big data techniques could infer a person's identity with ease.[181]

B. Outside of HIPAA and HITECH: Patient Generated and mHealth

As explained in the previous part, HIPAA, as amended by HITECH and regulations thereunder, applies to EHRs that are held by covered entities and their business associates. It does not apply to providers of health record systems directly to the consumer, nor does it apply to employer provided personal health records not connected to an

Research Exceptionalism Is Bad Medicine, 24 Health Matrix 49, 59–60 (2014) (asserting that medical information should be a public resource).

[177] *See* Janardhanan, *supra* note 95, at 702.

[178] 45 C.F.R. § 164.512 (2015).

[179] 45 C.F.R. § 164.514 (2015).

[180] *See* McGraw & Leiter, *supra* note 32, at 144–45.

[181] *See* Paul Ohm, *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization*, 57 UCLA L. Rev. 1701, 1740 (2010) (discussing methods for re-identification and arguing that HIPAA does not serve its purpose as a result). *But see* Khaled El Emam et al., *A Systematic Review of Re-Identification Attacks on Health Data*, 6 PLoS One e28071 (2011), http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3229505/ (asserting that studies do not include large databases and few include health data; therefore conclusions cannot be drawn).

employer-provided health plan.[182] Thus, it is important to note that the data collected in these external systems and the "mHealth" movement of data collection are primarily located outside of regulations to protect patient information.[183]

More broadly, however, the FTC will bring charges against a health portal or personal health record entity that engages in misrepresentations about its use of health information, or fails to reasonably implement security measures for the protection of that information.[184] In December 2013 the FTC settled a case against Accretive Health due to a security failure that led to the loss of a laptop containing twenty million data points relating to over 20,000 patients.[185] Accretive did not institute procedures to protect personal health information, a violation of section 5 of the FTC Act,[186] because it did not limit access and because it did not delete health information when it was no longer needed for billing purposes.[187]

Regulatory oversight of health data could occur tangentially by means of the Federal Communications Commission and Food and Drug Administration treatment of health product safety, though not directly related to the issues of big data and reform discussed herein.[188] Overall,

[182] *See* Deven McGraw, *Privacy and Health Information Technology*, 37 J.L. Med. & Ethics 123, 127 (2009) (noting that a group of eight of the largest employers who created electronic health platforms are not covered).

[183] *See* Helm & Georgatos, *supra* note 141, at 156 (noting that most health and wellness apps are not covered by HIPAA).

[184] *See* Lesley Fair, *When a Data Oops Becomes an Uh-oh*, FTC: Business Blog (Dec. 31, 2013, 1:15 PM), http://www.ftc.gov/news-events/blogs/business-blog/2013/12/when-data-oops-becomes-uh-oh. The FTC also enforces breach notification rules for personal health records outside of HIPAA. *See* 16 C.F.R. § 318.1 (2015).

[185] *See* Press Release, Fed. Trade Comm'n, Accretive Health Settles FTC Charges That It Failed to Adequately Protect Consumers' Personal Information (Dec. 31, 2013), http://www.ftc.gov/news-events/press-releases/2013/12/accretive-health-settles-ftc-charges-it-failed-adequately-protect [hereinafter Accretive Health Press Release].

[186] 15 U.S.C. § 45 (2012).

[187] Accretive Health Press Release, *supra* note 185.

[188] *See* Cortez, *supra* note 141, at 1200–16.

"federal regulators are sympathetic, not hostile, to mobile health"[189] and are "creating a very sympathetic regulatory environment."[190]

C. Nondiscrimination

The abominable history of discrimination in the provision of health care in the United States includes segregation, refusal to treat, and unequal treatment.[191] Title VI of the Civil Rights Act of 1964[192] addressed differential medical treatment by prohibiting discrimination in any "program or activity" supported by federal financial assistance, including health provision by the federal government, such as Medicare and Medicaid.[193] As a result, Title VI led to the desegregation of hospitals.[194] However, individual rights to sue under Title VI are limited to intentional discrimination, and only administrative actions are available for disparate impact discrimination.[195] A lack of Title VI enforcement more broadly has "cast a long shadow contributing to continuing racial segregation in America's health care delivery system."[196] In addition, studies show a continued history of differential treatment for specific medical conditions and a widening gap between mortality rates for Caucasian and African American citizens,[197] despite the adoption of national health programs that focused on these issues.[198] Gender and ethnicity

[189] *Id.* at 1217.

[190] *Id.* at 1214.

[191] *See* Watson, *supra* note 18, at 860.

[192] Pub. L. No. 88-352, tit. VI, 78 Stat. 241, 252 (1964) (codified as amended at 42 U.S.C. §§ 2000d–d-7 (2012)).

[193] Watson, *supra* note 18, at 861.

[194] *Id.* at 864.

[195] *See* Alexander v. Sandoval, 532 U.S. 275, 281, 293 (2001) (holding individuals may sue only for intentional discrimination). For a review of Title VI's history, see Watson, *supra* note 18, at 862–70.

[196] *See* Watson, *supra* note 18, at 866.

[197] *See* Yearby, *supra* note 21, at 1291–94.

[198] See, for example, the Healthy People program, first begun in 1979 and most recently updated for the adoption of goals set for 2020. *History & Development of Healthy People*, HEALTHYPEOPLE.GOV, https://www.healthypeople.gov/2020/about/History-and-Development-of-Healthy-People (last visited Dec 29, 2015).

differences, which are not related to other factors, continue to affect health outcomes.[199] This relationship is "stark and very troubling" as it "run[s] the fault lines of . . . persistent patterns of discrimination and disadvantage in society."[200]

More recent laws, the Genetic Information Nondiscrimination Act of 2008 (GINA),[201] and section 1557 of the ACA, also affect the use of data for health analytics.[202] GINA prohibits discrimination in employment and insurance based on genetic information.[203] In particular, health insurers may not require genetic testing as a precondition to obtaining insurance, and they may not increase premiums based on genetic information.[204] Employers generally may not acquire or use genetic information for broadly defined workplace decisions such as hiring, firing, or conditions of employment.[205] Neither may employers purchase such information about an employee or family member.[206]

Section 1557 of the ACA extended broad civil rights protections to health programs receiving any federal assistance or managed under a federal executive agency.[207] Covered entities include not only those who receive direct financial benefit, but also those who receive "credits, subsidies, or contracts of insurance."[208] ACOs, health insurance exchanges, as well as providers who accept Medicare or Medicaid payments, are covered by section 1557.[209] Entities that accept federal financial

[199] *See* Bobinski, *supra* note 14, at 365–68.

[200] *Id.* at 368.

[201] Pub. L. No. 110-233, 122 Stat. 881 (2008).

[202] ACA § 1557(a); 42 U.S.C. § 18116(a) (2015).

[203] See Ifeoma Ajunwa, *Genetic Testing Meets Big Data: Tort and Contract Law Issues*, 75 Ohio St. L.J. 1225, 1239–42 (2014) (limited application to insurers and employers).

[204] *Id.*

[205] *Id.*

[206] *Id.*

[207] *See* Daryll C. Dykes, *Health Injustice and Justice in Health: The Role of Law and Public Policy in Generating, Perpetuating, and Responding to Racial and Ethnic Health Disparities Before and After the Affordable Care Act*, 41 Wm. Mitchell L. Rev. 1129, 1199 (2015).

[208] ACA § 1557(a); 42 U.S.C. § 18116(a) (2015).

[209] *See* Watson, *supra* note 18, at 872–76.

incentives to adopt health technology systems would be covered under this section, at least until the incentives expire. Described as the "first Federal civil rights law to prohibit sex discrimination in health care,"[210] section 1557 prohibits discrimination based on race, color, national origin, age or disability and refers to Title VI of the Civil Rights Act, Title IX of the Education Amendments of 1972,[211] the Age Discrimination Act of 1975,[212] and the Rehabilitation Act of 1973[213] in order to define its scope and methods of enforcement.[214]

Section 1557 has also been described as a type of disability rights statute, based on the prediction that its nondiscrimination provisions will lead to more effective care for more complex health needs and a lessening of the concomitant burden on the overall lives of persons with disabilities.[215] Thus, the outcomes of adopting a policy of nondiscrimination support broader health-care goals. Whether laws effectively protect patients' privacy and prevent discrimination within the big data health-care dynamic created by the ACA is the topic of the following part.

IV. Toward Healthy Data Analytics

Policy makers are relying on big data and predictive analytics to solve a multitude of vexing health-care problems,[216] and the ACA requires the

[210] *See Section 1557 of the Patient Protection and Affordable Care Act*, U.S. Dep't Health & Hum. Servs., http://www.hhs.gov/ocr/civilrights/understanding/section1557/ (last visited Dec. 29, 2015).

[211] Pub. L. No. 92-318, 86 Stat. 235 (1972) (codified as amended at 20 U.S.C. §§ 1681–1688 (2012)).

[212] Pub. L. No. 94-135, tit. III, § 302, 89 Stat. 713, 728 (1975) (codified as amended at 42 U.S.C. §§ 6101–6107 (2012)).

[213] Pub. L. No. 93-112, 87 Stat. 355 (1973) (codified as amended at 29 U.S.C. §§ 701–714 (2012)).

[214] *See Fact Sheet: Nondiscrimination in Health Programs and Activities Proposed Rule: Section 1557 of the Affordable Care Act*, U.S. Dep't Health & Hum. Servs., http://www.hhs.gov/civil-rights/for-individuals/section-1557/summary/index.html#_ftn1 (last visited Jan. 8, 2016).

[215] *See* Jessica L. Roberts, *Health Law as Disability Rights Law*, 97 Minn. L. Rev. 1963, 1998–2000 (2013).

[216] *See supra* Part III.

expanded collection of information in order to award benefits for efficient quality health care, impose costs for falling short of data driven metrics, and identify health disparities among populations.[217] On the journey to achieve this vision, however, three problematic health-care data areas become intertwined: data-dependent policy, commercialization of big health data, and health disparities and data. First, policy overreliance on data to solve the problems of health care can lead to harm to personal privacy and autonomy and to an imbalance between healthy outcomes, financial incentives, and penalties based on data driven results. This is one of the contributing factors leading to the second problem. Patient information from health and nonhealth sources, exacerbated by mHealth applications outside of HIPAA protections, is becoming commercialized, thereby blurring the goals between individual health care and profit-driven interests. Finally, the reform goal of healthy outcomes can become skewed by using data that is not focused on eliminating the underlying causes of health disparities and that can create stigmatizing surveillance. Addressing these issues would strengthen the use of data for the purpose of efficiently delivering health care and promoting patient-centered care. The following analysis critically discusses the current state of each of these three areas, highlighting the ways that they are interrelated, and then suggests alternate or additional methods for addressing these weaknesses, so that data analytics in health care are used in healthy ways, both literally and figuratively.

A. Policy: Data Dependence and Health Information Privacy

Broad policy dependence on health data for reform is a major driver of commercial interests in health care and provides the underlying basis for financial incentives for healthy behaviors based on data driven metrics. However, a “data utopian viewpoint”[218] of problem solving can result in unintended consequences.[219] The combination of collection of

[217] *See supra* Part II.B.

[218] *See* Lupton, *supra* note 112, at 856 (trust and reliance on data to provide answers to varied and complex issues).

[219] *See* K. Krasnow Waterman & Paula J. Bruening, *Big Data Analytics: Risks and Responsibilities*, 4 Int’l Data Privacy L. 89, 94 (2014) (users of predictive analytics “must guard against overreach and consider the cumulative effect of analytics and Big Data”).

sensitive, personal health data from a wide variety of sources, both within and outside the clinical setting, and the application of predictive analytics to control cost and to provide health-care decisions, creates potential threats to individual privacy, autonomy, and fairness. It can even affect the decision about whether an individual is offered the option for a particular health treatment.[220] Furthermore, collection and use of data to identify disparities and social determinants of health, especially as data collection expands and mHealth tools spread, could negatively impact individual privacy and autonomy when the data are used for personal surveillance.

Health monitoring could have extraordinary benefits. However the *mis*uses could also impose personal harms as "[t]he privacy implications of individuated big data analysis are profound," and predictive modeling may turn into "dehumanizing 'data determinism.'"[221] Policy statements that support expansive use of data without fully recognizing and addressing the potential harms spur the use of data in this way. For example, contrary to the assumption made in the Big Data Report, that receiving an unwarranted call from a doctor is equivalent to receiving a bad movie recommendation, a patient's privacy is essential to the provision of proper health care. Repeated or offensive interventions can create real harms to the patient provider relationship. Studies have shown, for example, that patients who believe that their information will not be protected from disclosure are less likely to share important health information with their doctors.[222] Based on the social angst caused by the oft

[220] *Id.*

[221] Nicolas P. Terry, *Big Data Proxies and Health Privacy Exceptionalism*, 24 HEALTH MATRIX 65, 79 (2014).

[222] *See* Israel T. Agaku et al., *Concern About Security and Privacy, and Perceived Control Over Collection and Use of Health Information Are Related to Withholding of Health Information from Healthcare Providers*, 21 J. AM. MED. INFORMATICS ASS'N 374, 375 (2014) (reporting that more than twelve percent of those surveyed had withheld health information due to concerns, and smokers were more likely to withhold information); Ea Mulligan & Annette Braunack-Mayer, *Why Protect Confidentiality in Health Records? A Review of Research Evidence*, 28 AUSTRALIAN HEALTH REV. 48 (2004) (noting that although there are differences between country studies, patients react to confidentiality concerns by withholding information, giving inaccurate information, or not seeking treatment); Mark A. Rothstein, *The Hippocratic Bargain and Health Information Technology*, 38 J.L. MED. & ETHICS 7, 9 (2010) (stating that protecting health privacy is a "matter of public health" related to the privacy of sensitive information).

repeated Target example, where the store successfully predicted a young woman's pregnancy before her father knew,[223] it is reasonable to assume that a doctor who greeted a patient with predictions from unknown health data analytics would invoke similar privacy violation distress.

1. Weakness of Current Policy

General big data policy could play a powerful role in protecting individual health information and preserving privacy. However, in the year following issuance of the Big Data Report and the PCAST Report, the Office of Science and Technology cosponsored a workshop on the discriminatory effects of big data, and the White House Council of Economic Advisors studied differential pricing by businesses based on big data algorithms.[224] On February 5, 2015, an Interim Report was issued as a one-year follow up to the big data policy reports.[225] Health data was only mentioned to tout its benefits and to describe a new plan to quickly enroll one million people into a research cohort to move forward the study of personalized cures for cancer.[226] Although the Interim Report mentions statutes in need of updating,[227] none of these included health data rules or regulations. The report states that security and privacy are "paramount" concerns, and that "rigorous privacy protections" would be adopted based on prior lessons learned from other federal agencies and reports.[228] An FTC Data Broker Report recommendation[229] for legislation to protect consumer privacy was

[223] The Target example has been reported widely in the media and used as an example in many articles. *See, e.g.*, PCAST Report, *supra* note 75, at 12.

[224] White House, Big Data: Seizing Opportunities, Preserving Values, Interim Progress Report 6–7 (2015), https://www.whitehouse.gov/sites/default/files/docs/20150204_Big_Data_Seizing_Opportunities_Preserving_Values_Memo.pdf [hereinafter Big Data Interim Progress Report].

[225] *See* John Podesta, *Big Data and Privacy: 1 Year Out*, White House Blog (Feb. 5, 2015, 9:29 AM), https://www.whitehouse.gov/blog/2015/02/05/big-data-and-privacy-1-year-out.

[226] Big Data Interim Progress Report, *supra* note 224, at 11.

[227] *Id.* at 2–4 (noting that legislation or amendments proposed or forthcoming include the Consumer Bill of Rights, Student Digital Privacy Act, and the Personal Data Notification & Protection Act).

[228] *Id.* at 11.

[229] *See* FTC Data Broker Report, *supra* note 105, at vii.

referenced but not explicitly supported, and no mention was made of health-related information that is increasingly collected and aggregated outside the health-care system or personal information that can be fed back to health-care providers to create predictive patient scores.[230]

In January 2016, the FTC's report entitled "Big Data: A Tool for Inclusion or Exclusion?" mentions big data in health care to exemplify the benefits while remaining silent about its potential harms in an unchecked big data environment. Health data is mentioned significantly twice, to describe the benefits of big data for 'precision medicine' and to describe how big data can "[p]rovide specialized healthcare to underserved communities."[231] Precision medicine is described as beneficial not only for treatment, but for its ability to build upon data in order to prevent readmissions and to lower health-care costs.[232] Second, big data in health care is noted for increasing health care for underserved communities, an example being an IBM program that built a database for oncology diagnosis and treatment, thereby providing a resource for "rural and low income" areas that do not have appropriate doctors.[233] This section of the report shows once again that big data is being envisioned in large part for reducing costs, based on collecting "richer and more complete data,"[234] and seen as the answer to the expansion of health treatments. But richer data is a poor substitute for remedying the lack of doctors, and is an attenuated approach to ameliorate the conditions of poverty; overdependence on its power could perpetuate differential treatment based on geographical location and economic status. The report mentions only GINA,[235] not HIPAA or HITECH, as a

[230] *See* Jane Sarasohn-Kahn, Here's Looking at You: How Personal Health Information Is Being Tracked and Used 5, 8–9 (2014), http://www.chcf.org/~/media/MEDIA%20LIBRARY%20Files/PDF/PDF%20H/PDF%20HeresLookingPersonalHealthInfo.pdf (noting that data that individuals are not consciously aware are being used in different ways has been called "dark data").

[231] *Id.*

[232] *Id.*

[233] *Id.*

[234] *Id.*

[235] *Id.* at 18.

law to be considered when using big data for health care, and it only suggests that users of big data analytics should "[r]emember"[236] that correlation does not equal causation, and that it "may be worthwhile to have human oversight"[237] when using big data for important health decisions.[238]

Articulation of clear goals in policy documents that prioritize strategies and task appropriate agencies to adopt tactics to protect privacy, patient choice, and autonomy is essential. Health information is uniquely different, and the vague reference to both lessons learned from other agencies and the data utopianism of previous policy documents is minimal support for a policy to protect the privacy of health information. If privacy protection is to be rigorous, then as a matter of policy it must begin before health data programs are launched. It must be a precursor rather than an "effort."[239] If HHS policies were robust, this could largely compensate for the absence of a more focused high-level policy. Unfortunately, a review of the HHS's strategic plan for 2014–2018[240] does not disclose a robust policy for health data privacy in the era of big data.

The HHS plan is organized into four strategic goals: strengthening health care, advancing scientific knowledge and innovation, advancing the health and safety of individuals, and creating efficient, transparent and effective programs.[241] The most relevant policy statement is found in a subcategory of the strategic goal to strengthen health care, called an objective, which seeks to "improve health care and population health

[236] *Id.* at 32.

[237] *Id.*

[238] *Id.*

[239] *See* Jane Sarasohn-Kahn, *supra* note 230, at 10 (need system design for patient control); *see also* Melissa Healy, *Big Data, Meet Big Money: NIH Funds Centers to Crunch Health Data*, L.A. Times (Oct. 9, 2014, 5:47 PM), http://www.latimes.com/science/sciencenow/la-sci-sn-big-data-money-20141009-story.html ("[C]hallenges to be worked out . . . [include] how researchers can share data gleaned from electronic medical records without compromising the privacy of patients.").

[240] *HHS Strategic Plan*, U.S. Dep't Health & Hum. Servs., http://www.hhs.gov/about/strategic-plan/index.html (last visited Dec. 29, 2015).

[241] *Id.*

through meaningful use of health information technology."[242] The narrative states without elaboration that "a strong health IT infrastructure can help ensure patients' privacy. . . ."[243] The objectives are supplemented with strategies and performance goals, which provide in pertinent part that they will "work to ensure privacy and security of electronic health information," and will encourage vendors to incorporate security features, connected with privacy by design.[244] The commitment to privacy is therefore general and unsupported by any performance goals, and not discussed separately from security.[245] The impression left is that privacy is being protected by strong technical measures, rather than constituting a policy priority and fundamental value.

2. Federal Health IT Strategic Plan

The Federal Health IT Strategic Plan 2015–2020 was issued in draft form by the ONC in February, 2015 ("ONC Draft Plan")[246] and in final form on September 15, 2015 ("ONC Final Plan").[247] While the ONC Draft Plan addressed individual privacy only in general ways, the ONC Final Plan shows a disheartening lack of attention to the important individual privacy and civil rights issues that are inherent in a national health information collection and aggregation system. A comparison of

[242] *Strategic Goal 1: Strengthen Health Care*, U.S. Dep't Health & Hum. Servs., http://www.hhs.gov/about/strategic-plan/strategic-goal-1/index.html (last visited Dec. 29, 2015).

[243] *Id.*

[244] *Id.*

[245] "Security by design" is a term generally used by technical professionals to mean building security into a product or system rather than adding it later. Privacy by design is a term generally used by privacy professionals to mean building in privacy in the same way. See the distinctions and commonalities described in Ann Cavoukian & Marc Chanliau, Privacy and Security by Design: A Convergence of Paradigms 2, 13 (2013), https://www.ipc.on.ca/site_documents/PbDBook-From-Rhetoric-to-Reality-ch8.pdf.

[246] *Health IT Strategic Planning*, HealthIT.gov (Feb. 26, 2015), http://www.healthit.gov/policy-researchers-implementers/health-it-strategic-planning.

[247] *See* Karen B. DeSalvo et al., *ONC Publishes Final Federal Health IT Strategic Plan 2015–2020*, Health IT Buzz (Sept. 21, 2015, 12:31 pm), https://www.healthit.gov/buzz-blog/uncategorized/federal-health-it-strategic-plan-2015-2020/.

these documents reflects an escalating emphasis on sharing individual health information as an overarching policy.

The graphic on the front of the ONC Draft Plan consists of three connected circles containing the words Collect, Share and Use,[248] a visual cue that effectively communicates the intent for the future of health information. Protection of health information is not one of the circles, but patient privacy and choice is listed as one of the Federal Health IT Principles, both within and outside the health system.[249] Under the goal of sharing health information, Objective 2C states that "[a]ligning with the HHS' Secretary's Strategic Initiative highlighted in the *HHS Strategic Plan 2014–2018* federal actions seek to protect patients' health information, as it is electronically stored and shared and their privacy rights."[250] Interpretation of the sentence structure of this statement might lead one to infer that privacy rights were simply an add-on to the goal of improving the security of health information, but at least it is noted. Similarly, both Objectives 5A and 5B make explicit links between data, HIT, and privacy, in part by identifying a three-year goal of "[a]dvanc[ing] science and knowledge in creating and using sensors, mobile technology, medical devices, and assistive technologies that enable users to quantify and use personal health information *while protecting their privacy*."[251] Strategies to achieve this outcome are vague, but acknowledged. One example is pushing health information onto mobile devices and social networking platforms "*while protecting the privacy of the information*"[252]

The first instance of a potential policy change toward privacy protection is in the introductory letter from the National Coordinator. In the

[248] U.S. Dep't. Health & Hum. Services, Federal Health IT Strategic Plan 2015–2020, at 1, http://www.healthit.gov/sites/default/files/federal-healthIT-strategic-plan-2014.pdf.

[249] The principle states, "Respect individual preferences. Person-centered care embraces the value of the individual inside and outside the health system, where all entities honor individuals' privacy, needs, values, and choices regarding their information, health, and care." *Id.* at 7.

[250] *Id.* at 16.

[251] *Id.* at 26 (emphasis added).

[252] *Id.* (emphasis added). Objective 5A also specifies "[i]ncrease the number, timeliness, quality, and usability of federal health and other relevant data sets available for public use while protecting privacy," but it does not provide further details about how to protect privacy. *Id.* at 25.

ONC Draft Plan, the coordinator's letter notes the goal to "prioritize increasing the electronic collection and sharing of health information *while protecting individual privacy*."[253] The final letter from the coordinator makes no reference to privacy, while "signal[ling]. . . efforts . . . to include new sources of information and ways to disseminate knowledge quickly, securely and effectively."[254] The final letter further emphasizes increased collection of health information from such new technologies as individual mobile phones[255] and a broader definition of what should be contained in a health record, noting the value of collecting individual social and economic information as well as traditional physical health information.[256]

A list of principles for health IT precedes the ONC Final Plan. It explicitly includes the protection of individual privacy and autonomy,[257] but these specific principles are not always found in the related, specific four goals of the strategic plan. The four goals are to advance person-centered and self-managed health; transform health-care delivery and community health; foster research, scientific knowledge and innovation; and enhance the national HIT infrastructure.[258] The principle of individual privacy and autonomy for HIT is included in the "[b]e person-centered"[259] principle, but it is completely absent from the strategic goal to advance person-centered health care.[260] Another strategy is to "[s]upport health IT policies that make available products that securely integrate self-generated health information, self-reported outcomes, and genomic information into an individual's longitudinal care records and self-care and wellness technologies."[261] It fails to identify privacy issues associated with the increase in more widely collected information

[253] *Id*. at 2 (emphasis added).

[254] U.S. Dep't. Health & Hum. Servs., Federal Health IT Strategic Plan 2015-2010, at 4, https://www.healthit.gov/sites/default/files/9-5-federalhealthitstratplanfinal_0.pdf.

[255] *Id*. at 10.

[256] *Id*. at 11.

[257] *Id*. at 8.

[258] *Id*. at 6.

[259] *Id*. at 8.

[260] *Id*. at 34–35 (though not mentioning security and choice).

[261] *Id*. at 35.

outside the traditional health-care environment. Similar strategies, while resulting in particular benefits, can nevertheless negatively impact privacy and create new risks for individuals. Yet the ONC Final Plan is silent about the potential privacy impacts and risks of such practices as tracking patient location,[262] using predictive analytics for public health,[263] and the general extensive expansion of health data collection.

Goal four, to promote the HIT infrastructure contains specific subgoals for privacy. The ONC Final Plan states an intent to "[a]im toward privacy and security-related policies" that address an increase in health information sharing, "foster" regulations and "publish guidance" for "high-level principles" that "advance trust" and enable interoperability.[264] The language of provisions under Goal 4A is problematic because it only vaguely supports privacy goals in a discretionary way while emphasizing health information exchange that can pose increased risks to individual privacy and autonomy.

In comparison, Goal 4B holds promise. Strategies include continued attention to privacy protection under HIPAA, the certification of IT products that are privacy and security compliant, and the enforcement of privacy protections against HIPAA non-covered entities.[265] One stated goal is to seek technical and uniform solutions to implement personal privacy choices but only "when those choices are required."[266] There is no assumption of the value of individual choice over disclosure. Additional strategies under this goal relate to cybersecurity.

Viewing privacy and health equality within the entire context of the ONC Final Plan can lead to the interpretation that privacy is a secondary consideration at best, rather than a part of a holistic approach to the concerns of individuals about their personal, and more broadly collected, information that will be fed into an expanding health information system. In sum, the ONC Final Plan has the potential to serve as a strong policy vehicle, but it inconsistently recognizes and integrates privacy goals, vaguely stated, into the future use of health data and

[262] *Id.* at 38.

[263] *Id.*

[264] *Id.* at 42.

[265] *Id.* at 43.

[266] *Id.*

analytics. For the promise of this plan to be realized for purposes of health data policy, privacy concerns must be expressed in relationship to each relevant goal, and threats and challenges of data and personal privacy and autonomy must be explicitly addressed. Indeed, public comments filed by an array of diverse interest groups pointed out the need for and importance of privacy considerations that are so addressed. Planned Parenthood requested ONC to "spearhead a meaningful and thorough process around protecting patient privacy and confidentiality," especially with respect to collection from different devices potentially used in a health exchange.[267] The American Academy of Pediatrics singled out privacy and confidentiality concerns for adolescents because present electronic health systems cannot address this important component, without which "adolescents will likely forego seeking needed health care, especially for reproductive health, substance abuse, or mental health concerns."[268] Ascension Information Services, the "largest non-profit health system in the United States," supported the goal of aggregating data from multiple sources, but "firmly believe[d] that a patient's data belongs to the patient," and "urge[d] ONC to provide more specific information on how it sees both the federal government and the private sector advancing this goal."[269] The ONC Final Plan ignores these concerns.

Policies reflected in these strategic plans are the drivers of how data will be used in the provision of health care. It is essential that policy makers at every level explicitly promote the values of individual privacy within a data driven world, as public comments confirm. The ONC Final Plan takes steps toward this goal, but the public interest demands

[267] Letter from Dana Singiser, Vice President of Public Policy and Government Relations, Planned Parenthood Fed'n of Am., to Karen DeSalvo, Nat'l Coordinator for Health Info. Tech. (Feb. 6, 2015), http://www.healthit.gov/sites/default/files/comments_upload/comments-hit_strategic_plan_2015-2020_finalsubmission.pdf.

[268] Letter from Sandra G. Hassink, President of Am. Acad. of Pediatrics, to Karen DeSalvo, Nat'l Coordinator for Health Info. Tech. (Feb. 6, 2015), http://www.healthit.gov/sites/default/files/comments_upload/aap_comments_for_federal_health_it_strategic_plan_02_06_15_0.pdf.

[269] Letter from Mark D. Barner, Senior Vice President & CIO, Ascension, President & CEO, Ascension Info. Servs., to Karen DeSalvo, Nat'l Coordinator for Health Info. Tech. (Feb. 6, 2015), https://www.healthit.gov/sites/default/files/comments_upload/ascension-federal_health_it_strategy_comment_2_6_15_ais_template.pdf.

a more intentional, specific embrace of those actions and outcomes that will make a difference.

B. Commercial Entities and Health Data[270]

Commercial entities play a prominent role in the realignment and "unmooring"[271] of historical modes of health care by ACA reform. They also hold essential information technology and data analytics expertise for its implementation. The realignment has been described as a devolution from strong normative and ethical medical and physician accountability for care, where "[c]orporations are deliberately kept weak,"[272] to a system where "new technologies [will] create substantial opportunities for private entrepreneurship," and "well-being will be bought and sold in a commercial marketplace."[273] Commercial entities outside of the traditional health-care field are becoming adept at accessing sensitive health information and nonprotected health information in ways that can commoditize patient information as the basis for for-profit products.[274]

Business associates can legally aggregate data across health-care sources that covered entities cannot,[275] and under certain circumstances such associates may use de-identified information received from covered

[270] This section primarily contemplates the use of health information outside of HIPAA regulations.

[271] *See* William M. Sage & Kelley McIlhattan, *Upstream Health Law*, 42 J.L. Med. & Ethics 535, 536 (2014).

[272] *Id.* at 537.

[273] *Id.* at 536. The suggested differences include a reformulation of dyadic care, modified physician control, hospital walls, third party payment, and physician-extending technology. *Id.* at 536–40.

[274] *See* Lupton, *supra* note 112, at 856; Wharam & Weiner, *supra* note 44, at e83 ("These stakeholders will dip into the burgeoning reservoir of health data to make predictions that advance their organizational agendas and financial well-being.")

[275] *See* Reece Hirsch & Heather Deixler, *HIPAA Business Associates and Health-Care Big Data: Big Promise, Little Guidance*, Bloomberg BNA (Feb. 21, 2014), http://www.bna.com/hipaa-business-associates-and-health-care-big-data-big-promise-little-guidance/.

entities in commercial products.[276] One law firm advises business associates to "where possible, . . . negotiate the ability to use PHI [personal health information], or at least de-identified health information, received from the covered entity to further their monetization initiatives."[277] It also has been argued that regulations should be expanded so that commercial entities can use health information more broadly than currently allowed.[278] Finally, there can be pressure to re-identify data, for example to intervene in patient behaviors, thereby completing the circle of commercialization of health information.[279] Individual health information provided by covered entities to business associates can be incorporated within commercial health analytic products that are then sold back to the health-care provider for providing health treatments at the point of care. In essence, the patient's information is commoditized and sold back to a physician.[280]

[276] For example, a contract clause between a covered entity and business associate provides as follows:

> Business Associate may use, sell, rent and otherwise disseminate the Inflexxion Behavioral Health Tools Data in aggregated, de-identified form for any purpose, or in the form of a Limited Data Set for Authorized Purposes, or in the form of analyses of such de-identified or Limited Data Set information, in its sole discretion, and that Covered Entity will not be entitled to any compensation for such use of Inflexxion Behavioral Health Tools Data.

INFLEXXION, BUSINESS ASSOCIATE/LIMITED DATA SET USE AGREEMENT 2, https://www.asimvconnect.com/pdf/Business%20Associate%20Agreement.pdf.

[277] FOLEY & LARDNER, LLP, TAPPING INTO THE BIG VALUE OF HEALTH CARE BIG DATA: TOP LEGAL AND REGULATORY CONSIDERATIONS ON THE PATH TO MONETIZATION 9 (2015), https://www.foley.com/files/Publication/b5702375-940f-4379-ba5f-f2e885088780/Presentation/PublicationAttachment/b74426c3-097c-4381-8366-3cfd3a0b852e/Monetization%20of%20Data%20-White%20Paper.pdf.

[278] For example, McKinsey & Company argue that regulations for the use of health information should be revisited in order to "encourage data sharing" and that "data sharing could be made the default, rather than the exception." *See* GROVES ET AL., *supra* note 103, at 13.

[279] *See* Terry, *supra* note 36, at 407–09 (noting that re-identification in the hands of business associates may include the argument that consent for further use is implied); *see also* Ohm, *supra* note 181 (discussing re-identification methodology); N. Nina Zivanovic, *Medical Information as a Hot Commodity: The Need for Stronger Protection of Patient Health Information*, 19 INTELL. PROP. L. BULL. 183, 190–91 (2015) (asserting that re-identification of de-identified information outside of HIPAA requires a legislative response to protect patient rights).

[280] *See* INFLEXXION, *supra* note 276.

Additionally, commercial products that use detailed health and social data, analytics, and algorithms[281] to create predictive models for patient care and health operation efficiencies have been dubbed "black boxes" due to "the use of opaque computational models to make decisions related to health care."[282] The unknown and unknowable bases for health and efficiency predictions, based on very large data sets compiled from a myriad of different sources, raises concern for equality of treatment. Monetization of health information, profit-driven business models, and the propertization of algorithms raise policy questions for individual privacy and public goods.[283]

Thus, in one vision of a future health data world, the push for healthcare reform and quantifiable outcomes will result in business entities, especially data brokers, putting commercial interests above individual health equality and autonomy. Aggregation of health data will occur from across so many sources and platforms, especially due to mHealth, that meaningful patient privacy will be lost. To achieve an alternative vision, where patient privacy coexists with data analysis, and there is a public private partnership for the promotion of health and equality of services, industry standards will be necessary. The standards could be regulatory, self-regulatory, or a combination of both. From the commercial health data analytics standpoint, the FTC may be the entity best poised to lead such an effort.

1. The Federal Trade Commission

The FTC has been a leader for many years in protecting individual privacy rights during tumultuous times of technology changes.[284] Indeed,

[281] For a further discussion, see *supra* Part III.B.

[282] W. Nicholson Price II, *Black-Box Medicine*, 28 Harv. J.L. & Tech. 419, 421 (2015); *see also* Frank Pasquale, The Black Box Society: The Secret Algorithms That Control Money and Information 3 (2015) (using the term black box as a metaphor for the opaqueness of both how data are collected and the consequences of their use).

[283] Terry states the concern more strongly, as "those who aggregate and mine this data neither view their informational assets as public goods held on trust nor seem particularly interested in protecting the privacy of their data subjects." *See* Terry, *supra* note 36, at 389.

[284] The FTC's leadership is well known, but beyond the scope of this article to chronicle. For background information and a survey of how FTC authority may affect corporate actions, see Kenneth A. Bamberger & Deirdre K. Mulligan, *Privacy on the Books and on the Ground*, 63 Stan. L. Rev. 247, 273–76 (2011).

two recent, broader FTC reports, the Data Broker Report[285] and the Internet of Things staff report,[286] are relevant to health data, indicating the agency's willingness to tackle thorny issues. In 2014, the FTC Data Broker Report described the industry[287] and identified potential harms[288] from an overlapping and extensive data environment that offers little transparency.[289]

Although data collection and analysis could bring personalized benefits to consumers, the Data Broker Report concludes with a unanimous recommendation for legislation to protect consumers from an opaque industry that threatens to impose harms.[290] Industry best practices also were recommended,[291] as was legislation for three types of data products: marketing, risk mitigation, and people search.[292] While people searching might touch health information tangentially, marketing products and risk mitigation are most closely related to health data.

When data brokers use consumer information for marketing purposes, especially sensitive data like personal health data, the FTC recommends that individuals have a right to see what data a broker possesses, and have the right to opt out of marketing based on that data. This framework would "allow consumers to control uses of the data about which they care the most."[293] Furthermore, the report recommends that legislation "requir[e] that consumer-facing sources obtain

[285] *See* FTC Data Broker Report, *supra* note 105.

[286] *See* FTC IoT, *supra* note 130.

[287] *See* FTC Data Broker Report, *supra* note 105, at 46–47 (summarizing industry characteristics).

[288] *Id*. at 48–49 (noting that problems arising from the use of health data could negatively affect a person's trust, create inferences that are unassailable by consumers, and create risks due to lengthy retention of the data).

[289] *Id*. at 49. ("Data brokers acquire a vast array of detailed and specific information about consumers; analyze it to make inferences about consumers, some of which may be considered quite sensitive; and share the information with clients in a range of industries. Much of this activity takes place without consumers' knowledge.").

[290] *Id*.

[291] *Id*. at 54–56.

[292] *Id*. at 50–54.

[293] *Id*. at 52.

consumers' affirmative express consent before collecting and sharing such [health] information with data brokers."[294]

The legislative recommendation does not touch on the most problematic issues involved with health data. Instead, these issues are addressed by voluntary best practices.[295] Identified best practices include adoption of privacy by design, the protection of children and youth, and auditing of practices to avoid discriminatory results and the improper use of data downstream.[296] Privacy by design principles involves approaching the entire design of a system or product from a positive sum, proactive viewpoint, and using privacy as the default choice in system design, including transparency that is "subject to independent verification."[297]

The FTC Internet of Things staff report treads softly on an emerging technology, preferring not to recommend any new legislative protections so as not to negatively affect industry developments.[298] It did, however, pledge to enforce current laws, work with industry and stakeholders, and to develop best practices in order to promote privacy for consumers.[299] Although a subsequent FTC workshop reviewed the environment of consumer created health data, no report emerged as a result, and more recent commentary from FTC sources refers to the conclusions in the Internet of Things

[294] *Id*. Though beyond the scope of this discussion, the report also describes an electronic portal where consumers could manage the information held by data brokers in one place. *Id*. at 50–51.

[295] *Id*. at 54–56. But two members believed that legislation was warranted. *Id*. at 56 n.108.

[296] *Id*. at 55–56.

[297] Intelligent Assistive Tech. & Sys. Lab & Info. & Privacy Comm, Ontario, Canada, Sensors and In-Home Collection of Health Data: A Privacy by Design Approach 17 (2010), https://www.ipc.on.ca/images/Resources/pbd-sensor-in-home.pdf. For a description of privacy by design applied to home health sensors, and a list of the general privacy by design principles, see *id*. at 16–17.

[298] *See* FTC IoT, *supra* note 130, at 48–49. However, it did continue to recommend new overall privacy legislation that could apply in general to data collection by this technology. *See id*. at 50–52.

[299] *Id*. at 53.

report when discussing best practices[300] for the use of consumer health data.[301]

The weakness of both the FTC Data Broker and the Internet of Things reports is that they do not explicitly address the complicated health environment where policy encourages the sharing of information from both inside and outside the health-care system to create longitudinal data about day-to-day personal lifestyles and (un)healthy behaviors. As commercial entities collect and analyze health-related data outside of traditional regulation, yet become intertwined with health-care data and entities, each will have potentially varied interests in the data, from commercial use and profit to health-care efficiency and health research.

The FTC Internet of Things Report builds upon the past approach of fair information practice principles (FIPPs)[302] that while still relevant do not address the difficulty of integrated health data, eHealth devices, and services. Placing the burden of controlling this complex data relationship upon individuals should not be the main approach. If the FTC's conclusion that more robust legislation is premature, even for sensitive health data, then the development of a self-regulatory, specific health information privacy framework could be the most effective step to protect health privacy at this point in time. The National Institute for Standards and Technology (NIST) recently released a proposed privacy risk management framework for federal information systems.[303] Although it is more general in nature, the framework could offer an approach for defining and implementing health privacy in a complex data environment.

[300] *See* Cora Han, *Using Consumer Health Data: Some Considerations for Companies*, FTC Business Blog (Apr. 28, 2015, 9:52 AM), https://www.ftc.gov/news-events/blogs/business-blog/2015/04/using-consumer-health-data-some-considerations-companies. See *supra* Part III.A.1 for further discussion of the workshop.

[301] *See* Cora Han, *Using Consumer Health Data?*, FTC Business Blog (Apr. 27, 2015, 9:32 AM), https://www.ftc.gov/news-events/blogs/business-blog/2015/04/using-consumer-health-data.

[302] *See* FTC IoT, *supra* note 130, at 19–20. The Fair Information Practice Principles (FIPPs) are: notice, choice, access, accuracy, data minimization, security, and accountability. *Id*. at 19.

[303] Nat'l Inst. for Standards & Tech., Privacy Risk Management for Federal Information Systems (2015), http://csrc.nist.gov/publications/drafts/nistir-8062/nistir_8062_draft.pdf [hereinafter NIST PRM].

2. A NIST-like Approach

In May 2015, NIST released "Privacy Risk Management for Federal Information Systems" (PRM)[304] that, while written for federal systems, can be adopted voluntarily by private entities.[305] The novel framework laid out a privacy engineering approach to assessing and managing risks to privacy, defined as "a collection of methods to support the mitigation of risks to individuals arising from the processing of their personal information within information systems."[306] In essence, the framework is a way that any entity can "first identify its goals and obligations for privacy protection, assess its systems against these governing requirements, prioritize mitigation mechanisms, and monitor for changes."[307]

While FIPPs are important principles and are reflected in the privacy engineering objectives, the present and future data intense world allows information to be used in ways that makes the FIPPs like "forcing a square peg into a round hole,"[308] whereas what is sought by technologists is a "repeatable and measurable method[] for identifying, prioritizing, and mitigating privacy problems."[309] The NIST view of FIPPs resonates with the privacy concerns related to ways for individuals to exert control over health information, mobile health data and big data, and reflects the PCAST Report view that consumer control is too difficult, and the entity collecting and using data should bear the responsibility for implementing privacy preferences.[310] Privacy engineering

[304] *Id.*

[305] *See, e.g.*, NOKIA, PRIVACY ENGINEERING AND ASSURANCE: THE EMERGING ENGINEERING DISCIPLINE FOR IMPLEMENTING PRIVACY BY DESIGN 1 (2014), http://www.w3.org/2014/privacyws/pp/Hirsch.pdf ("Privacy Engineering and Assurance is the engineering methodology to bridge the gap between laws and principles and technologies.").

[306] NIST PRM, *supra* note 303, at 4 n.1 (noting first though that privacy engineering is a new discipline without an accepted definition).

[307] *Id.* at 15.

[308] *Id.* at 7. In comparison, FIPPs are somewhat subjective, contextual, and applied inconsistently. *Id.* at 8.

[309] *Id.* at 9.

[310] *See* PCAST REPORT, *supra* note 75, at 44.

objectives adopted in the PRM are predictability, manageability, and disassociability.[311] These objectives are basic building blocks for building privacy-responsive systems and serve to "bridg[e] the gap between an [entity's] . . . goals for privacy and their manifestation in information systems."[312]

Predictability is the "reliable belief about what is occurring with personal information" that is "core to building trust and enabling self-determination."[313] The description of the predictability objective explains that the use of data and corresponding results should not surprise individuals, and risk analysis should identify if it will do so.[314] This objective is particularly relevant for health information privacy, and essential for the provision of effective health care. A person must be able to trust the health-care provider and be able to exert her autonomy in order to be willing to share personal health and lifestyle information. In addition, the objective illustrates the two-way importance of patient privacy and autonomy. If health analytics are to serve an increasingly important role in the efficient delivery of health care, then the data upon which the analysis and algorithms are based must be accurate and dependable, a function of an individual's trust to share the information. There are two corollaries to the predictability principle. First, if unpredictable results occur, then system operators should take steps to halt the use or access to those results, with de-identification, for example.[315] Second, because information system monitoring is undertaken to permit remedial actions if unpredictable changes arise, the predictability objective allows for innovation and new uses.[316] As new products and new methods for health-care analytics develop, there are likely to be miscalculations or misapplications of that technology, but if the monitoring and course

[311] *See* NIST PRM, *supra* note 303, at 18.

[312] *Id.*

[313] *Id.*

[314] *Id.* at 19 ("[P]redictability facilitates the maintenance of stable, trusted relationships between information systems and individuals and the capability for individuals' self-determination, while enabling operators to continue to innovate and provide better services.").

[315] *Id.*

[316] *Id.*

correction occur, then trust may be preserved and support an innovative environment.

Manageability—the second PRM objective—does not refer to the individual's control over information, but rather to the fine-grained administration of that system in order to "implement key FIPPs, including maintaining data quality and integrity, achieving data minimization, and implementing individuals' privacy preferences."[317] In applying the PRM to health care, manageability should refer more specifically to HIPAA and all other health-related legal requirements.

The third PRM objective—disassociability—is defined as "actively protect[ing] . . . an individual's identity or associated activities from unnecessary exposure."[318] This objective resonates with particular strength to entities' use of personal health and behavioral information. The potential for re-identification of an individual amidst the fusion of information from many sources and the inference of health information from data that are outside the health system are two areas for application of this objective. A health-care covered entity is now only required under one provision to lack the actual knowledge that data could be re-identified,[319] rather than a proactive approach taken by the disassociability objective. Outside of the health-care setting, using predictive analytics in ways that would infer health information or status about an individual and making it available in more widespread ways would be avoided under this objective.[320]

In addition to privacy engineering objectives, the PRM includes a privacy risk equation, defined as the multiple of the likelihood of a problematic data action and the impact of that action.[321] The risk analysis requires an organization to identify data actions that can be problematic

[317] *Id.*

[318] *Id.* at 20. Data actions are defined as the processing of personal information that "can include, but is not limited to . . . collection, retention, logging, generation, transformation, disclosure, transfer, and disposal." *Id.* at 22.

[319] *See supra* Part IV.A.

[320] See the Target example *supra* note 223 and accompanying text.

[321] *See* NIST PRM, *supra* note 303, at 22.

and the specific privacy issues associated therewith, and to determine how to mitigate or accept the risk.[322] Problematic data actions are ones that would cause an adverse impact, or problems, for individuals.[323] These problems are categorized as loss of self-determination, discrimination, loss of trust, and economic loss.[324] But the risk equation is not an end to itself. It is part of a continual process in which problems are identified, assessed, prioritized, and mitigated. If a data operation has low risk for occurring and a low harmful impact, then the entity may decide to simply monitor it, whereas if an operation has a high risk for significant harm, it would be a high priority to address. Within the health environment, the same process could be applied to the anticipated benefits of data analytics, assessing the risk and likelihood of privacy and discriminatory harms as well as the anticipated benefits and likelihood of outcomes.

More specifically, potential harms identified within these categories resonate with the reasons for protecting the privacy of health data: loss of autonomy, exclusion, stigmatization, power imbalance, and loss of trust.[325] While problems affect the health-care relationship from a privacy standpoint, they are also closely tied to the problems of health disparities and health equality. If not carefully implemented, health surveillance and interventions may impose stigmas on populations, exclusion from certain treatments, and loss of trust in the health-care system when data and predictive analytics create differential treatment.

In sum, the NIST privacy engineering framework could be modified for use for health data management.[326] Furthermore, a specific health privacy-engineering framework could be created to integrate the unique situation of health data and the provision of health care with the

[322] *Id*. at 24.

[323] *Id*. at 29.

[324] *Id*. at 55 ("For example, sensitive information such as health data or criminal records or merely accessing certain services such as food stamps or unemployment benefits may attach to individuals creating inferences about them.").

[325] *Id*.

[326] ONC already utilizes the Cybersecurity Risk Framework developed by NIST, which "aligns with the HIPAA risk assessment." Kathryn Marchesini et al., *Building a Culture of Health IT Privacy and Security*, HEALTH IT BUZZ (Oct. 23, 2014, 3:03 PM), http://www.healthit.gov/buzz-blog/privacy-and-security-of-ehrs/building-culture-health-privacy-security/.

increasing use of data collection and analytics from differing sources. The emphasis on monitoring for risks and privacy harms to individuals, manageability for the ability to make fine-grained control of data, predictable use of data that includes broader considerations than minimal personal identification, and the fundamental preference for disassociation, responds to the world of health data in which incentives promote the sharing of information from different and ubiquitous sources. Application of a health-privacy risk framework would facilitate analysis that would consider the context of the health data use and the nature of the harms. Importantly, the continual monitoring of privacy risks would help address the evolving uses of different communications and medical technology that is becoming increasingly person centered. While the framework would not eliminate entirely the need for considering legislation for regulating privacy and the mHealth and health data marketplace, it would more closely align the attributes of patient–physician confidentiality and medical ethical obligations with the commercial sector.

C. Health Disparities and Data

As discussed previously in Part II.B, the ACA requirement to collect data about race, sex, and other personal information is motivated by the policy and goal to eliminate health disparities. The incentive structure of the ACA is supportive of that goal, insofar as it is based on promoting healthy behaviors and interventions to prevent illness. However, success is measured by data and numerical goals, independent of the means used to reach those goals, or their implications. Policies that favor the use of data to achieve results overlook the problems created by overreliance on data and predictive analysis. The problems may result in the loss of autonomy, discrimination, and stigmatization while not addressing the underlying determinants of health. These are the types of harms that could be assessed and remedied in what may be more accurately called a health-privacy and equality framework, following a process that would consider these individual harms in the application of analytics.

Health monitoring and surveillance could have extraordinary benefits for individuals and result in the promotion of healthy outcomes. However these uses could also impose personal harms because "[t]he privacy implications of individuated big data analysis are profound," and predictive modeling may turn into "dehumanizing 'data determinism.'"[327] Expressed otherwise, data-driven surveillance and interventions could be considered a type of data paternalism. Negative individual impact can occur due to the effects of paternalism and a perceived loss of personal decision making.[328] However, "[o]n empirical grounds, there can be no question that people who exercise the greatest degree of individual autonomy also enjoy the best health. Conversely, people with the least amount of autonomy—the least amount of control over their work conditions or other major life circumstances—have the poorest health."[329] Furthermore, a patient is less likely to trust a system that is perceived to be based on surveillance and coercion, and that intrudes upon a sense of privacy and autonomy, as it has been noted that "[t]he divide between surveillance as an exercise of power and surveillance for public goals can sometimes be blurry."[330] While a complete discussion of public health, surveillance, and paternalism is beyond the scope of this discussion, it is important to note the relevance of health data to these issues and an individual's sense of lack of control.

In a health system where "research has suggested that doctors misinterpret population data in their patient assessments, allowing racial stereotypes to influence their perceptions, leading them to automatically

[327] *See* Terry, *supra* note 221, at 79.

[328] *See* David Adam Friedman, *Public Health Regulation and the Limits of Paternalism*, 46 CONN. L. REV. 1687, 1769 (2014) (discussing different forms of paternalism and a recommendation to tailor their uses depending on the circumstances; "[r]egulators should pursue all solutions open to them-but they should do so with a cost-benefit rationalization that includes the likelihood that paternalism will present an obstacle to implementing the solution."); *cf*. Wendy E. Parmet, *Beyond Paternalism: Rethinking the Limits of Public Health Law*, 46 CONN. L. REV. 1771, 1790 (2014) ("Public health laws should not be seen as the edict of a disembodied policymaker seeking to benefit an unwilling public.").

[329] David R. Buchanan, *Autonomy, Paternalism, and Justice: Ethical Priorities in Public Health*, 98 AM. J. PUB. HEALTH 15, 17 (2008).

[330] Pasquale, *supra* note 5, at 770. While not based on big data, a study by Ralph L. Keeney, *Personal Decisions Are the Leading Cause of Death*, 56 OPERATIONS RES. 1335 (2008), exemplifies how failing to take into account the social determinants of health can create harmful stereotypes and lead to more intense surveillance. The study used statistics to show that personal decisions caused approximately fifty percent of deaths in the United States; these included the decision to smoke, the decision to be obese (i.e., to eat, or not exercise, according to the author), the decision to take drugs, and the like. Without considering the determinants of health such as stress, access to healthy foods, and the like, the study could be used to stigmatize persons who have made these choices.

associate diseases arising from behavioral choices—such as drug abuse and obesity—with blacks, while being less prompt in accurately identifying these conditions in white patients,"[331] institutionalized discriminatory treatment poses a risk to quality health-care analytics. Furthermore, studies have shown that even perceived discrimination is a contributing variable to negative health outcomes.[332]

For example, Part III.B described the laudable joint project between Carilion and IBM to predict which patients would develop heart problems, and to take steps to intervene and prevent the illness. A predictive model will be created using data mined from nonstructured comments of health-care workers. But it is likely that data from patient charts will contain biases and prejudices of doctors and health-care professionals (endemic to health care, even if unintentional) and therefore will be incorporated into the predictive model. In order to avoid this result, the data would need to be scrubbed of the discriminatory data that feed into the model. While it is unknown whether these problems were considered in this project, it is a difficult task because there is little or no way to identify which records are subject to bias, absent words that are obvious discriminatory triggers.

"[D]ata overflow and opaqueness of algorithms"[333] can make it difficult if not impossible to identify discrimination. In addition, while disparate impact analysis is available to prove discrimination under Title VI, there is no private cause of action, therefore limiting available remedies.[334]

One might argue that the result in the above example is not harmful to individuals. The misdiagnosed patient would receive reminders and interventions that would still lead to a healthier lifestyle. This is not the whole picture, however. The bias in the model could result in patients

[331] Matthew, *supra* note 19, at 5.

[332] *See* Social and Behavioral Domains, *supra* note 11, at 33.

[333] Kelsey Finch & Omer Tene, *Welcome to the Metropticon: Protecting Privacy in a Hyperconnected Town*, 41 Fordham Urb. L.J. 1581, 1603 (2014).

[334] *See* Matthew, *supra* note 19, at 54–58 (arguing that the law should be amended to provide for a private cause of action). It has been argued by one author, however, that the ACA implies such a remedy. *See* Sarah G. Steege, Note, Finding a Cure in the Courts: A Private Right of Action for Disparate Impact in Health Care, 16 Mich. J. Race & L. 439 (2011).

who were *not* included for intervention, and one would never really know how interventions could have helped those persons. In fact, even if the algorithm was not based on physician commentary it is still likely it could be discriminatory because historical data about treatments would incorporate a record of discrimination based on characteristics such as sex, race, or ethnicity.

Lastly, interventions based on predictive algorithms, which are intended to encourage healthy behaviors and decrease health-care costs by granting economic incentives or penalties based upon outcomes, may correlate social or demographic factors with behaviors that could serve to stigmatize groups unfairly. A dystopian future could include high-income patients being sought out for health-care treatments and commercially targeted for their value to marketers, while low-income patients are targeted for behavior modification and for the purpose of decreasing their use of the health system. Data flows from technically sophisticated and educated users may originate from a voluntary lifestyle monitor into a patient-centered care model while aggregated data flows may send underserved populations into a health-care responsibility and surveillance model, thereby perpetuating disparate treatment in health care.[335] More broadly, the inclusion of healthy behaviors or activity incentives in employer-provided health plans means that the implications from predictive health analytics could "lead to stigmatization and discrimination in the workplace and elsewhere if the public perceives certain 'races' as more diseased or more difficult to treat than others."[336]

[335] *See* Hoppin, *supra* note 37, at 1982–986, 1991 (arguing that public health surveillance should have to meet a strict scrutiny standard for health issues such as obesity, and that it cannot); Lindsay F. Wiley, *Access to Health Care as an Incentive for Healthy Behavior? An Assessment of the Affordable Care Act's Personal Responsibility for Wellness Reforms*, 11 IND. HEALTH L. REV. 635, 640–41 (2014) ("Personal responsibility reforms reflect cultural biases that exaggerate the extent to which ill health is attributable to the personal failings of unhealthy individuals and . . . serve as a political distraction from less punitive measures aimed at making our communities, workplaces, schools, and marketplaces more conducive to healthy living.").

[336] *See* Sharona Hoffman, *"Racially-Tailored" Medicine Unraveled*, 55 AM. U. L. REV. 395, 398 (2005). For further discussion of the effect of behavioral health programs, see Jessica L. Roberts, *"Healthism": A Critique of the Antidiscrimination Approach to Health Insurance and Health-Care Reform*, 2012 U. ILL. L. REV. 1159 (2013); Jessica L. Roberts, *Healthism and the Law of Employment Discrimination*, 99 IOWA L. REV. 571 (2014).

A NIST-like risk management framework lends itself to the implementation of steps to avoid risks of discrimination and employing mitigation strategies while not precluding interventions. The analysis would require consideration of the likelihood of the problem occurring, such as discrimination or stigmatization, and the magnitude of the harm. Policy that recognizes the potential harms is also important and can be instrumental in spurring implementation of the risk management framework.

The importance of providing health care without discrimination is too important, however, to limit it to a voluntary approach. An amendment should be made to section 1557 of the ACA allowing for review of data driven decision-making tools used in health care in order to eliminate discriminatory impact, applied to health information, wherever found.[337] The statutory amendment should grant HHS the power to adopt rules for proving disparate impact. Furthermore, ACA incentives should be revised in order to motivate health-care companies to implement a system of privacy protections that are similar to the NIST privacy engineering framework by considering the harms and mediating them.

Conclusion

The convergence of health-care reform incentives, ubiquitous data collection, blurred boundaries between relevant health and behavioral information, growing commercial interests in health data, and thorny health disparities, raises serious questions for big data in health care. This combination threatens the core of individual privacy and autonomy, and could result in differential treatment based on group membership such as race or sex. Harm is incurred by numerous factors, including an overreliance on data to produce economic savings; the unexamined collection and use of health-related data that is expanding to lifestyle information beyond that traditionally collected for health

[337] Auditing of data analytic products has been suggested in other contexts. *See* Amy J. Schmitz, *Secret Consumer Scores and Segmentations: Separating "Haves" from "Have-Nots"*, 2014 Mich. St. L. Rev. 1411, 1470–72 (describing auditing techniques for data analytics used for credit scoring). Requiring transparency of algorithms is a related method for providing insight. *See* Omer Tene & Jules Polonetsky, *Big Data for All: Privacy and User Control in the Age of Analytics*, 11 Nw. J. Tech. & Intell. Prop. 239, 270–72 (2013)..

reasons; and the broad, unaudited use of data analytics. Addressing these problems will require a multipronged approach, beginning with policy leadership that recognizes the existence of the problems and the importance of addressing them with specific strategies. A risk management framework can assist in multiple ways by bringing both commercial and health-care entities to a review of their data practices and products, an assessment of the problematic actions that they may produce for individuals, and a plan to remediate consequences. Lastly, legislative action is called for to address potential discriminatory applications of data analytics and the resulting surveillance, and to modify incentives. In summary, the ACA's goals of improving health care and eliminating health disparities can be achieved by means of healthy data policy and practices, so that they do not lead to unhealthy consequences.