
BIG DATA IN HEALTH CARE— PREDICTING YOUR FUTURE HEALTH

KRISTINA FUNAHASHI*

TABLE OF CONTENTS

INTRODUCTION	356
I. PART I.....	357
A. PREDICTIVE ANALYTICS PRIMER.....	357
B. (“BIG”) DATA SOURCES.....	359
1. Clinical Data	359
2. Claims and Cost Data	360
3. Pharmaceutical R&D Data	360
4. Nonhealthcare Data Sources.....	361
C. PREDICTIVE HEALTH ANALYTICS EXAMPLES	362
D. POTENTIAL RISKS POSED BY PREDICTIVE ANALYTICS IN HEALTHCARE	364
1. Psychological Harms & The Duty to Disclose	365
2. Discrimination Based on Future Health Status.....	366
i. Healthcare Insurers	366
ii. Healthcare Providers.....	367
iii. Employers.....	367
II. PART II.....	368
A. CURRENT PRIVACY & NONDISCRIMINATION LAWS.....	368
1. Data Privacy & Protection: The Health Insurance Portability and Accountability Act	368

* Executive Articles Editor, *Southern California Law Review*, Volume 94; J.D. Candidate 2021, University of Southern California Gould School of Law; B.A. Organismic and Evolutionary Biology 2014, Harvard University. I would like to thank Professor Alexander M. Capron for his invaluable guidance and insights during the drafting of this Note. I would also like to thank the Southern California Law Review Staff for their incredibly detailed and diligent assistance throughout the editing process. Last but far from least, a heartfelt thank you to my grandfather, Jerry D. Wu, M.D., and my parents, Lenora and Ted Funahashi, for their unwavering encouragement, love, and support.

2. Health Insurance: The Patient Protection and Affordable Care Act (ACA).....	371
3. Nondiscrimination Laws: The ADA and GINA.....	373
i. The Americans with Disabilities Act.....	373
ii. The Genetic Information Nondiscrimination Act.....	375
B. GAP ANALYSIS.....	378
1. HIPAA.....	378
2. The ACA.....	378
3. The ADA.....	379
4. GINA.....	379
5. Nonlegal Protections.....	379
C. SUMMARY.....	380
III. PART III.....	380
A. SIMILARITIES BETWEEN GENETIC TESTING & PREDICTIVE ANALYTICS.....	381
B. DIFFERENCES BETWEEN GENETIC TESTING & PREDICTIVE ANALYTICS.....	383
1. Diagnostic Screening Versus System-Enhancing Tool....	383
2. Impact on Relatives.....	384
C. RECOMMENDATIONS FOR FUTURE PREDICTIVE HEALTH ANALYTICS REGULATION.....	385
1. Antidiscrimination Legislation.....	386
2. Disclosure of Predictive Health Analytics Outcomes.....	387
3. Post-ACA Considerations.....	388
4. Life Insurers and Other Stakeholders.....	389
CONCLUSION.....	390

INTRODUCTION

Predictive analytics—a branch of data analysis that generates predictions about future outcomes through the power of computers to process large amounts of data using statistical modeling and machine learning—is increasingly applied in health care. While it has the potential to improve patient health and lower health care costs, the ability to peer into people’s future health status has also raised significant concerns about privacy and patient self-determination. Part I of this Note explains predictive analytics and machine learning in healthcare; it discusses data sources (which may not all be medical records) and examines several predictive analytics models. It concludes by assessing the risks posed by predictive health analytics, including psychological harms to patients and discrimination by healthcare insurers, healthcare providers, and employers. Part II summarizes existing federal data privacy and nondiscrimination

legislation relevant to healthcare information in order to assess where the law leaves gaps regarding the regulation of predictive health data. By comparing predictive health analytics with genetic testing—another method of predicting an individual’s risk of disease where laws have been enacted to protect perceived “misuses” of test results—Part III reaches conclusions about how the law could treat the use of predictive health analytics and makes recommendations about future protections for patients.

I. PART I

A. PREDICTIVE ANALYTICS PRIMER

Predictive analytics is the use of big data and electronic algorithms to forecast the likelihood of future events based on historical data.¹ Predictive analytics draws its power from a range of analytical methods and techniques.² This includes data mining, or the exploration and analysis of large data sets to discover meaningful patterns, and machine learning, which enables computers to detect and use patterns in data sets, to predict future outcomes.³ The use of predictive analytics has accelerated in numerous industries in the past decade with the emergence and expansion of real-time, electronic data sets for which traditional data-processing tools have been insufficient.⁴ Today, predictive analytics has helped to build numerous commercial and financial tools, such as Amazon’s recommendation system for online shopping, efficient pricing systems in the stock market, and customer profitability predictions in banking.⁵

The healthcare industry has lagged behind others in implementing predictive analytic models due to the complexities of healthcare and the generally slower rate at which it has adopted electronic technologies for data management.⁶ However, the advent of electronic health records (“EHRs”) and healthcare reform in the early 2000s, which is bringing an increasing

1. See I. Glenn Cohen, Ruben Amarasingham, Anand Shah, Bin Xie & Bernard Lo, *The Legal and Ethical Concerns that Arise from Using Complex Predictive Analytics in Health Care*, 33 HEALTH AFF. 1139, 1139 (2014).

2. See DAVID CROCKETT, RYAN JOHNSON & BRIAN ELIASON, HEALTH CATALYST, WHAT IS DATA MINING IN HEALTHCARE? 1–2 (2017), <https://www.healthcatalyst.com/wp-content/uploads/2014/06/What-is-data-mining-in-healthcare.pdf> [<https://perma.cc/7AVB-GFW2>].

3. *Id.*

4. Cohen et al., *supra* note 1, at 1140.

5. See David W. Bates, Suchi Saria, Lucila Ohno-Machado, Anand Shah & Gabriel Escobar, *Big Data in Health Care: Using Analytics to Identify and Manage High-Risk and High-Cost Patients*, 33 HEALTH AFF. 1123, 1128 (2014); CROCKETT ET AL., *supra* note 2, at 1 (“[Data mining] serves similar use cases in telecom, manufacturing, the automotive industry, higher education, life sciences, and more.”).

6. CROCKETT ET AL., *supra* note 2, at 2.

proportion of healthcare services under large systems, began a healthcare data “revolution” that set the foundation for predictive analytics in healthcare.⁷

The electronic storage of individuals’ health information can provide patients and providers with faster and more efficient information processing, retrieval, and communication. This information is also available to researchers, which enables them to develop complex computational techniques such as predictive analytics to understand the needs of patients—as individuals and as populations—with greater precision, even before such needs manifest in illness or disability.⁸ The Affordable Care Act of 2010 (“ACA”) further stimulated widespread use of electronic health data in healthcare.⁹ Two of the ACA’s objectives were to improve population health and decrease health disparities among populations. Meeting these objectives required accumulating a vast amount of data to build models used in assessing the relative effectiveness of various approaches to improving outcomes and cutting costs.¹⁰ Moreover, payment reforms such as “accountable care,” which aims to incentivize providers and health systems to deliver care more effectively and efficiently by making them accountable for those outcomes, and “bundling,” a payment approach in which healthcare providers deliver a set of services for a predefined price, have motivated healthcare organizations to improve the efficiency of their care.¹¹ Such incentives, in conjunction with widespread adoption and use of EHRs,¹² have led to increasing interest from healthcare insurers, payers, and providers in the use of predictive analytics to deliver better care while simultaneously reducing costs.¹³

7. Janine S. Hiller, *Healthy Predictions? Questions for Data Analytics in Health Care*, 53 AM. BUS. L.J. 251, 257 (2016).

8. Sharona Hoffman & Andy Podgurski, *In Sickness, Health, and Cyberspace: Protecting the Security of Electronic Private Health Information*, 48 B.C. L. REV. 331, 332 (2007).

9. Hiller, *supra* note 7, at 258.

10. *Id.* at 259.

11. Bates et al., *supra* note 5, at 1124.

12. More than 95% of hospitals possess an EHR system. See SONAL PARASRAMPURIA & JAWANNA HENRY, OFFICE OF THE NAT’L COORDINATOR FOR HEALTH INFO. TECH., ONC DATA BRIEF NO. 46: HOSPITALS’ USE OF ELECTRONIC HEALTH RECORDS DATA, 2015–2017 1 (2019), <https://www.healthit.gov/sites/default/files/page/2019-04/AHAEHRUseDataBrief.pdf> [<https://perma.cc/P5VQ-AC2L>].

13. Malay Gandhi & Teresa Wang, *The Future of Personalized Healthcare: Predictive Analytics*, ROCK HEALTH, <https://rockhealth.com/reports/predictive-analytics> [<https://perma.cc/2L8L-G6JF>] (investment in predictive analytics companies grew from \$201 million in 2011 to \$902 million in 2014).

B. (“BIG”) DATA SOURCES

Data is generated by almost everyone and everything that interfaces with the healthcare system.¹⁴ In this emerging era of big data and machine learning, predictive analytics models are able to use a variety of current and historical data sources. Traditional sources of healthcare data include clinical data and insurance claims data.¹⁵ However, data outside of the healthcare industry, such as consumer purchasing patterns and social media, are increasingly used to make predictions about an individuals’ future health.¹⁶ This Section will briefly describe several sources of data used to develop and power predictive analytics models in healthcare.

1. Clinical Data

Clinical data are owned by healthcare providers and are primarily collected in EHRs or electronic medical records (“EMRs”), which contain longitudinal records of patient health information generated by one or more encounters in any care delivery setting.¹⁷ EMRs are digital versions of the paper charts in the clinician’s office, while EHRs are broader and designed to focus on the total health of the patient, as well as built to share information with other healthcare providers, such as laboratories and specialists.¹⁸ EHRs and other medical records contain an enormous amount of health data.¹⁹ EHRs include patient demographics, progress notes, current problems, medications, vital signs, visit history, past medical history, immunizations, radiology reports, as well as laboratory and other test results.²⁰ Adoption of EHRs by physicians and health systems continues to increase, in large part due to the Medicare Meaningful Use program,²¹ which encourages providers to use certified EHRs and demonstrate “meaningful use” by providing incentive payments to providers that meet EHR utilization thresholds for a

14. See Nathan Cortez, *Substantiating Big Data in Health Care*, 14 *V/S*: J.L.P. 61, 67 (2017).

15. Cohen et al., *supra* note 1, at 1140–41.

16. *Id.*

17. See Office of the Nat’l Coordinator for Health Info. Tech., *What is an Electronic Health Record (EHR)?*, HEALTHIT.GOV, <https://www.healthit.gov/faq/what-electronic-health-record-ehr> [<https://perma.cc/UV7C-TJV3>].

18. See Peter Garrett & Joshua Seidman, *EMR vs. EHR—What is the Difference?*, HEALTHIT: BUZZ, <https://www.healthit.gov/buzz-blog/electronic-health-and-medical-records/emr-vs-ehr-difference> [<https://perma.cc/K2P5-26DR>]. However, the various EHR systems are not interoperable with each other.

19. For reference, Kaiser Permanente, an integrated payer-provider health network, maintains over forty-four petabytes of EHR data for its roughly 9.1 million patients, which is forty-four times more data than that contained in the entire Library of Congress. Cortez, *supra* note 14, at 67.

20. *Id.*

21. Meg Bryant, *EHR Data Increasingly Used to Inform Clinical Care Decisions*, HEALTHCARE.DIVE (Apr. 18, 2019), <https://www.healthcaredive.com/news/ehr-data-increasingly-used-to-inform-clinical-care-decisions/553002> [<https://perma.cc/4JFW-G8V7>].

number of objectives that can positively impact patient care. Eligible providers that do not participate in the program are subject to financial penalization via a reduction in their Medicare reimbursements.²²

2. Claims and Cost Data

Insurance claims and cost data are another sort of electronic record of health information. Claims and cost data are generally owned by the payer or insurer which covers the patient.²³ Claims data is generated whenever a healthcare encounter between an insured patient and physician, hospital, pharmacy, or other healthcare provider occurs. As compared to EHR data, claims data is broad and captures information from all doctors and providers caring for a patient, while EHR data captures only the portion of care administered by providers using their EHR system.²⁴ Claims data consists of billing codes that providers submit to payers, such as commercial insurance companies and government health plans (for example Medicare), and includes information such as diagnosis codes, procedure codes, and prescription information.²⁵ Example data sets include utilization of care and cost estimates.²⁶

3. Pharmaceutical R&D Data

Pharmaceutical research and development (“R&D”) data is another source of big data for predictive modeling. R&D data describes drugs’ therapeutic mechanisms of action, target behavior in the body, side effects, toxicity, and other information generated by research studies and clinical trials.²⁷ Such data is owned by the pharmaceutical companies or academic institutions that conduct the research.²⁸

22. Ctrs. for Medicare & Medicaid Servs., *CMS Finalizes Definition of Meaningful Use of Certified Electronic Health Records (EHR) Technology*, CMS.GOV (July 16, 2010), <https://www.cms.gov/newsroom/fact-sheets/cms-finalizes-definition-meaningful-use-certified-electronic-health-records-ehr-technology> [https://perma.cc/JM3Q-NDQT].

23. PETER GROVES, BASEL KAYYALI, DAVID KNOTT & STEVE VAN KUIKEN, *THE ‘BIG DATA’ REVOLUTION IN HEALTHCARE: ACCELERATING VALUE AND INNOVATION* 3–4 (2013), https://www.ghdonline.org/uploads/Big_Data_Revolution_in_health_care_2013_McKinsey_Report.pdf [https://perma.cc/8SHJ-GF6C].

24. JOHN WILSON & ADAM BOCK, *THE BENEFIT OF USING BOTH CLAIMS DATA AND ELECTRONIC MEDICAL RECORD DATA IN HEALTH CARE ANALYSIS* 2–4 (2012), <https://www.optum.com/content/dam/optum/resources/whitePapers/Benefits-of-using-both-claims-and-EMR-data-in-HC-analysis-White-Paper-ACS.pdf> [https://perma.cc/7EH6-SZSR].

25. *See id.* at 2.

26. *See* GROVES ET AL., *supra* note 23, at 4.

27. *See id.* at 3–4.

28. *Id.*

4. Nonhealthcare Data Sources

More recently, predictive health analytics models have drawn upon sources traditionally situated outside of healthcare, such as patient behaviors and preferences, retail purchase history, and data captured from fitness trackers.²⁹ Data from smartphones, wearables, and other remote monitoring sensors generate a vast amount of continuous data on individuals, as compared to episodic data generated by encounters with the healthcare system.³⁰ It is estimated that 60% of adults in the United States track their diet, weight, and exercise regimes by utilizing various types of mobile devices, like the Apple HealthKit.³¹ Consumer data such as shopping patterns, GPS location data, internet site visits, social media use, and other sources can also be used to make inferences about an individuals' health.³² Data miners and scientists now consider social media posts, blogs and status updates on social media to be within the realm of big data for healthcare due to the information's correlation with health outcomes.³³ The volume and variety of healthcare data sources continues to grow,³⁴ spurred by an ever-growing body of evidence that behavioral and environmental factors account for more than half of individuals' health outcomes.³⁵

Patient behavior data from trackers or websites is generally owned by the companies who provide the product or service or data brokers—companies that collect, aggregate, and sell information about consumers from public and private sources.³⁶ For example, data brokers aggregate information from public records, court records, and property records, with sources purchased or collected from commercial sources, such as purchase histories, social media sites, and web-browsing activities, then sell these data to interested parties.³⁷ Healthcare payers can purchase these data sets in order to learn about patient finances, purchasing preferences and other

29. *Id.* at 4.

30. Cortez, *supra* note 14, at 63–64.

31. Lia Steakley, *Harnessing Mobile Health Technologies to Transform Human Health*, STAN. MED.: SCOPE (Mar. 16, 2015), <http://scopeblog.stanford.edu/2015/03/16/harnessing-mobile-health-technologies-to-transform-human-health> [<https://perma.cc/824T-NNUV>].

32. Cortez, *supra* note 14, at 64–65.

33. Hiller, *supra* note 7, at 270.

34. The amount of healthcare data worldwide is expected to grow to 25,000 petabytes, about fifty times more than the 500 petabytes estimate in 2012. Cortez, *supra* note 14, at 64–65.

35. *Harnessing Nontraditional Data Sources*, OPTUM, <https://www.optum.com/resources/library/new-consumer-data.html> [<https://perma.cc/MJ8M-VVEY>].

36. Hiller, *supra* note 7, at 271–72.

37. Yael Grauer, *What Are 'Data Brokers,' and Why Are They Scooping Up Information About You?*, VICE: MOTHERBOARD (Mar. 27, 2018, 7:00 AM), https://www.vice.com/en_us/article/bjpx3w/what-are-data-brokers-and-how-to-stop-my-private-data-collection [<https://perma.cc/9FVJ-UEKL>].

characteristics from data brokers such as Acxiom, a company that has collected over 3,000 data points for nearly every consumer in the U.S.³⁸ Healthcare payers also own stand-alone data analytics divisions that perform data-broker functions, such as Optum, owned by the massive U.S. payer UnitedHealth Group.³⁹ Optum has collected socioeconomic, clinical, and cost data of 150 million Americans dating back to 1993, which it reports using to link patient outcomes to details like their education, net worth, and family structure.⁴⁰ The company also filed a patent application in 2016 to gather information from Facebook and Twitter, and link this to the person's clinical and payment information.⁴¹ Data brokers are not subject to U.S. medical privacy laws, such as HIPAA; this has generated concern from the healthcare industry and privacy advocates alike.⁴²

C. PREDICTIVE HEALTH ANALYTICS EXAMPLES

Early uses of predictive analytics in healthcare have focused on identifying patients at risk for serious complications or adverse clinical events in order to prevent them and efficiently allocate scarce resources.⁴³ For example, in 2014, IBM collaborated with Epic, an EHR provider, and the Carilion Clinic, a Virginia health system, to identify 8,500 patients at risk of heart failure in order to implement early intervention and better care.⁴⁴ The results were achieved through predictive modeling using Carilion's EHR data, including clinicians' notes and discharge documents using IBM's natural language processing technology to analyze and understand this information in the context of the EHR.⁴⁵ Patients in the pilot program identified as at risk for heart failure within one year were made prime

38. Hiller, *supra* note 7, at 271.

39. Marshall Allen, *Health Insurers Are Vacuuming Up Details About You – And It Could Raise Your Rates*, NPR (July 17, 2018, 5:00 AM), <https://www.npr.org/sections/health-shots/2018/07/17/629441555/health-insurers-are-vacuuming-up-details-about-you-and-it-could-raise-your-rates> [<https://perma.cc/ZBT6-GWZX>].

40. *Id.*

41. *Id.*

42. See generally Sharona Hoffman, *Big Data Analytics: What Can Go Wrong*, 15 IND. HEALTH L. REV. 227 (2018) (explaining challenges related to medical research, including privacy concerns).

43. *Id.*

44. See Hiller, *supra* note 7, at 279; Press Release, IBM Predictive Analytics to Detect Patients at Risk for Heart Failure (Feb. 19, 2014) [hereinafter IBM Press Release], <https://www-03.ibm.com/press/us/en/pressrelease/43231.wss> [<https://perma.cc/T5MQ-EEBK>].

45. See IBM Press Release, *supra* note 44. Prior to advanced analytic capabilities, early detection and prevention of heart failure has been challenging. IBM's natural language processing technology further enabled the use of "unstructured" data from physician notes and other documents, which enabled more complete and accurate information to be used as a data source for the model. *Id.*

candidates for care management and early interventions.⁴⁶ Predictive models based on insurance claims data have also been used to predict short-term mortality for elderly patients, classify cancer patients with different responses to chemotherapy, and predict the prognosis of patients receiving thoracic organ transplantation.⁴⁷

More recently, researchers have also been able to use predictive analytics to forecast the likelihood of certain diseases even further into the future, long before the diseases manifest themselves. Such models have been referred to as “long-term forecasts.”⁴⁸ For example, new research suggests that deep learning algorithms may be able to predict the onset of Alzheimer’s disease.⁴⁹ The researchers used thousands of positron-emission tomography (“PET”) images of individuals’ brains and other data from the Alzheimer’s Disease Neuroimaging Initiative to teach a deep learning algorithm to detect early signs of Alzheimer’s, on average more than six years in advance of diagnosis by a physician.⁵⁰ Such a model could assist clinicians by enabling them to identify signs of Alzheimer’s long before disease manifestation and improve opportunities for early treatment.⁵¹ Another study released in 2019 suggests that predictive analytics models may also be a valuable tool for predicting the risk of premature death attributable to chronic disease.⁵² The scientists found that machine learning models significantly improved the accuracy of prediction of premature all-cause mortality in middle-aged adults in the United Kingdom as compared to standard clinical methods.⁵³ The algorithm was able to analyze a vast and “holistic” array of individual

46. *Id.*

47. Kun-Hsing Yu, Andrew L. Beam & Isaac S. Kohane, *Artificial Intelligence in Healthcare*, 2 NATURE BIOMED. ENG’G 719, 726 (2018), <https://doi.org/10.1038/s41551-018-0305-z> [<https://perma.cc/SD45-R8P8>].

48. Sharon Hoffman, *What Genetic Testing Teaches About Predictive Health Analytics Regulation*, 98 N.C. L. Rev. 123, 124–27 (2019) [hereinafter Hoffman, *Genetic Testing & PHA Regulation*].

49. Catharine Paddock, *Alzheimer’s: Artificial Intelligence Predicts Onset*, MED. NEWS TODAY (Nov. 7, 2018), <https://www.medicalnewstoday.com/articles/323608.php#1> [<https://perma.cc/3Z4A-GVNR>].

50. *Id.* The model was also statistically more effective than radiologists at recognizing patients who would go on to have a subsequent diagnosis of Alzheimer’s disease. *Id.*

51. Although current interventions are unlikely to successfully treat Alzheimer’s, they focus primarily on helping people maintain mental function, manage behavioral symptoms, and slow down the progression of the disease. *Treatment of Alzheimer’s Disease*, NAT’L INST. ON AGING, <https://www.nia.nih.gov/health/how-alzheimers-disease-treated> [<https://perma.cc/362F-F6P4>].

52. Stephen F. Weng, Luis Vaz, Nadeem Qureshi & Joe Kai, *Prediction of Premature All-Cause Mortality: A Prospective General Population Cohort Study Comparing Machine-Learning and Standard Epidemiological Approaches*, 14 PLOS ONE, Mar. 27, 2019, at 1, 2, <https://doi.org/10.1371/journal.pone.0214365> [<https://perma.cc/9BT3-AYEF>].

53. *Id.* at 17.

clinical, demographic, lifestyle, and environmental risk factors, to achieve a degree of predictive accuracy beyond that of standard approaches.⁵⁴

While the aforementioned models were built primarily using data obtained from health systems, hospitals, insurers, and other healthcare providers, data from a variety of nonhealth sources can also be used. The health insurance industry recently joined forces with data brokers to collect large nonhealth-related data sets that include information such as race, education level, TV habits, and marital status to predict healthcare costs.⁵⁵ For example, the Carolinas HealthCare System combined a data set of individuals' purchases using store loyalty cards with de-identified data from a healthcare data broker in order to generate a patient "risk score" used to identify patients that would benefit from contact prior to experiencing any health problems.⁵⁶ Social media is also an increasingly common source of predictive analytics data and resulting models. For instance, Facebook recently developed machine learning algorithms to help identify people who may be at risk of suicide or self-harm to provide them with resources and services.⁵⁷ In order to train the model, Facebook data mined tens of thousands of Facebook posts reported by concerned friends.⁵⁸ Facebook now uses its algorithms to scan nearly every post on the platform to assess suicide risk, then passes the information along to law enforcement for wellness checks.⁵⁹

D. POTENTIAL RISKS POSED BY PREDICTIVE ANALYTICS IN HEALTHCARE

Predictive analytics models have the potential to be extremely valuable to healthcare providers, insurers, and researchers in order to improve patient outcomes and decrease healthcare costs. Instead of waiting for symptoms or other physical manifestations or for the results of specific tests, providers

54. *Id.*

55. Allen, *supra* note 39.

56. Hiller, *supra* note 7, at 279–80; Shannon Pettypiece & Jordan Robertson, *Hospitals Soon See Donuts-to-Cigarette Charges for Health*, BLOOMBERG (June 26, 2014, 9:35 AM), <http://www.bloomberg.com/news/articles/2014-06-26/hospitals-soon-see-donuts-to-cigarette-charges-for-health> [<https://perma.cc/CN25-75DE>].

57. Diana Kwon, *Can Facebook's Machine-Learning Algorithms Accurately Predict Suicide?*, SCI. AM. (Mar. 8, 2017), <https://www.scientificamerican.com/article/can-facebooks-machine-learning-algorithms-accurately-predict-suicide> [<https://perma.cc/6S3Z-PLXB>].

58. *Id.*

59. Benjamin Goggin, *Inside Facebook's Suicide Algorithm: Here's How the Company Uses Artificial Intelligence to Predict Your Mental State from Your Posts*, BUS. INSIDER (Jan. 6, 2019, 8:19 AM), <https://www.businessinsider.com/facebook-is-using-ai-to-try-to-predict-if-youre-suicidal-2018-12> [<https://perma.cc/S99D-JW99>]. Facebook has yet to release results regarding the accuracy of its predictive model to detect suicide risk. *Id.*

and researchers can use predictive analytics by relying on a wide range of existing data to detect individuals' disease risks and future health outcomes years before symptoms occur. However, this ability to predict individuals' future health outcomes also raises a number of concerns, including psychological harms to patients, privacy violations, and the risk of discrimination based on predictions of future health status. The collection and use of big data from sources outside of clinical, claims, and R&D data, such as data sets sold by data brokers aggregated from public records, social media, and other commercial sources, increases these concerns since nonhealthcare data lacks many of the protections afforded to healthcare data.⁶⁰

We turn first to an examination of the potential psychological harms predictive health analytics may create for patients, which also raises question about the extent to which providers' duty to disclose to patients in the course of treatment extends to the possible health outcomes generated by predictive analytic tools. We will then focus on the risk of discrimination by healthcare insurers, payers, providers, and employers if they become aware of individuals' predicted health outcomes.

1. Psychological Harms & The Duty to Disclose

The ability to predict an individual's future health is highly valuable to the healthcare system since it enables providers to improve treatment outcomes, enhance accuracy of diagnosis and treatment, and maximize scarce resources.⁶¹ However, providers have found, in conveying predictions based on genetic screening and diagnosis, that individuals who learn they are at risk for developing life-threatening or debilitating conditions are likely to find the news devastating.⁶² Such traumatic news may impact their ability to concentrate at work, their personal relationships, and their overall mental health.⁶³ Likewise, studies on the emotional impact of a dementia diagnosis have shown that individuals experience shame, discrimination, rejection, social isolation, and altered self-image.⁶⁴

Similar reactions occur from disclosure of "incidental findings," that is, results that "are outside the original purpose for which the test or procedure

60. Hiller, *supra* note 7, at 281.

61. *See infra* Part I.

62. Hoffman, *Genetic Testing & PHA Regulation*, *supra* note 48, at 140–41.

63. *Id.*

64. Faranak Aminzadeh, Anna Byszewski, Frank J. Molnar & Marg Eisner, *Emotional Impact of Dementia Diagnosis: Exploring Persons with Dementia and Caregivers' Perspectives*, 11 *AGING & MENTAL HEALTH* 281, 287 (2007).

was conducted.”⁶⁵ This has particular relevance in predictive health analytics. While some predictive models are based on set algorithms, many predictive models are based on unsupervised machine learning to seek patterns in data.⁶⁶ Rather than search for a particular result, a machine learning model may analyze a mass of data and identify complex patterns that conventional human analysis likely would not have discerned.⁶⁷ As a result, incidental findings are likely to be common, or even inherent, to the field.⁶⁸ As the field of predictive health analytics progresses, providers should address whether or not they have a duty to disclose to patients that their health system—or others who have access to patients’ data in the EHR—are running computer programs that can produce predictions about patients’ future health.

2. Discrimination Based on Future Health Status

In addition to potential psychological harms to patients, individuals also face the risk of discriminatory treatment in healthcare and employment as a result of predictive analytics results. This Section will focus on the risk of discrimination by healthcare insurers, providers, and employers on the basis of an individual’s predicted health outcomes.

i. Healthcare Insurers

Healthcare insurers have an interest in keeping their covered populations healthy, an aim which predictive analytics will help to achieve by allowing them to identify issues in their patients early and provide the services they need. But health insurers also have an interest in signing up healthy people and avoiding sick people in order to reduce payments for care.⁶⁹ Although the ACA prohibits insurers from denying individuals health insurance coverage or varying premiums based on pre-existing conditions,⁷⁰ patients’ personal information may still be used for marketing and to assess risks and determine the prices of certain plans. Insurers have a long history of attempting to avoid sick people and still may attempt to do so in subtle

65. PRESIDENTIAL COMM’N FOR THE STUDY OF BIOETHICAL ISSUES, ANTICIPATE AND COMMUNICATE: ETHICAL MANAGEMENT OF INCIDENTAL AND SECONDARY FINDINGS IN THE CLINICAL, RESEARCH, AND DIRECT-TO-CONSUMER CONTEXTS 2 (2013) [hereinafter INCIDENTAL FINDINGS].

66. KYLIE WATSON, DELOITTE, PREDICTIVE ANALYTICS IN HEALTHCARE: EMERGING VALUE AND RISKS 2 (2019).

67. *Id.* For example, a predictive model originally intended to detect early signs of Alzheimer’s disease may find that certain individuals are also at an increased risk for other, unrelated conditions.

68. See Hoffman, *Genetic Testing & PHA Regulation*, *supra* note 48, at 151–52.

69. Allen, *supra* note 39.

70. Patient Protection and Affordable Care Act (ACA) § 1201, 42 U.S.C. §§ 300gg to 300gg-7 (2018). The ACA will be discussed in further detail in the next section.

ways. For instance, as reported by Marshall Allen, one company established “its enrollment office on the third floor of a building without an elevator, so only healthy patients could make the trek to sign up.”⁷¹ Other companies may place certain drugs (for example HIV drugs) in their highest co-pay tiers, neglect to provide adequate information about which drugs they cover,⁷² or may outright exclude or limit certain types of providers for their networks in order to discourage certain patients from enrolling.⁷³

ii. Healthcare Providers

Similarly, shifts in the manner in which healthcare providers are paid, for example, toward capitated payments for a patient and away from payments based on the quantity of services provided, make patients who are likely to be and to remain healthy more attractive to medical providers because caring for such patients will be less expensive.⁷⁴ Healthcare providers may choose whether to accept particular new patients, and could exercise this discretion based on information about a person’s predicted health status in an effort to reduce costs.⁷⁵

iii. Employers

Employers are also likely to be interested in future health information about job candidates and current employees. Employers are incentivized to hire workers with maximum productivity and minimal health risks in order to minimize absenteeism and healthcare costs. Since most Americans are covered through a work-related insurance plan, employers have an incentive to avoid employees who will generate high health insurance costs.⁷⁶ Although the Americans with Disabilities Act (“ADA”) does not permit

71. Allen, *supra* note 39.

72. Although insurers who do so run the risk of penalization for failure to disclose such information during the enrollment period. Abigail English & Julie Lewis, *Privacy Protection in Billing and Health Insurance Communications*, 18 AMA J. ETHICS 279, 281 (2016).

73. Allen, *supra* note 39; Olga Khazan, *How Insurance Companies Still Discriminate Against the Sick*, ATLANTIC (Jan. 28, 2015), <https://www.theatlantic.com/health/archive/2015/01/how-insurance-companies-still-discriminate-against-the-sick/384908> [<https://perma.cc/CMC3-ZZUQ>].

74. Allen, *supra* note 39. In the United States, the only time a patient has a right to care is when that patient needs true emergency care in a hospital emergency room under the Emergency Medical Treatment and Labor Act (EMTALA) or under the “anti-dumping” law. *EMTALA Fact Sheet*, AM. C. EMERGENCY PHYSICIANS, <https://www.acep.org/life-as-a-physician/ethics--legal/emtala/emtala-fact-sheet> [<https://perma.cc/SSGZ-29QC>].

75. Trisha Torrey, *Doctors Firing or Dismissing Patients*, VERYWELL HEALTH (June 9, 2020), <https://www.verywellhealth.com/doctors-reject-difficult-patients-deny-medical-care-2615006> [<https://perma.cc/G5P3-XUV9>]. This is particularly concerning as physician shortages worsen and patient demand increases.

76. Sharona Hoffman, *Big Data and the Americans with Disabilities Act*, 68 HASTINGS L.J. 777, 779 (2017) [hereinafter Hoffman, *Big Data and the ADA*].

employers to conduct medical examinations or make medical inquiries during the application process,⁷⁷ they may still gain medical information about applicants by examining their ability to perform job-related tasks.⁷⁸ Employers can also conduct medical examinations after hiring an employee to determine whether the employee is able to continue performing responsibilities safely.⁷⁹ Some employers also require workers to authorize release of their medical records upon request.⁸⁰ Moreover, employers can obtain data on candidates and employees through social media and often base employment decisions on information discovered on sites such as Facebook and Twitter.⁸¹ Employers are thus likely to be particularly interested in predictive health analytics models that draw from both healthcare and nonhealthcare sources.

II. PART II

Looking to the future, as predictive analytics becomes capable of producing more detailed and reliable information about health outcomes, do current data privacy and nondiscrimination laws adequately protect patients? Or do gaps remain in protecting patients and allowing health analytics to be used in a way that provides a net benefit?

A. CURRENT PRIVACY & NONDISCRIMINATION LAWS

1. Data Privacy & Protection: The Health Insurance Portability and Accountability Act

The Health Insurance Portability and Accountability Act (“HIPAA”), which was signed into law in August 1996, was intended to improve the efficiency and effectiveness of the healthcare system.⁸² The statute gave the Department of Health and Human Services (“HHS”) authority to develop regulations to protect the privacy of medical records. After an extensive process, the HIPAA Privacy Rule (“Privacy Rule”) was promulgated in

77. 42 U.S.C. § 12112(d) (2018).

78. Hoffman, *Big Data and the ADA*, *supra* note 76, at 780.

79. *Id.*

80. *Id.* There are around twenty-five million compelled authorizations in the U.S. each year. Mark A. Rothstein, *Currents in Contemporary Bioethics: Access to Sensitive Information in Segmented Electronic Health Records*, 40 J.L. MED. & ETHICS 394, 397 (2012).

81. Hoffman, *Big Data and the ADA*, *supra* note 76, at 780.

82. U.S. Dep’t of Health & Human Servs., *HIPAA for Professionals*, HHS.GOV, <https://www.hhs.gov/hipaa/for-professionals/index.html> [<https://perma.cc/Q8M7-7W6Y>]. Privacy and Security rule compliance for small health plans was required on April 14, 2004, and April 20, 2006, respectively. *Id.*

2000,⁸³ followed by the HIPAA Security Rule (“Security Rule”) in 2003.⁸⁴

The Privacy Rule empowers the Office for Civil Rights (“OCR”), part of the HHS, “to penalize the unauthorized disclosure of ‘protected health information’ ” (“PHI”).⁸⁵ PHI is defined as “individually identifiable health information . . . [t]ransmitted by electronic media[,] [m]aintained in electronic media[,] or [t]ransmitted or maintained in any other form or medium,” by a HIPAA “covered entity,” which includes healthcare providers, health plans and insurers, and healthcare clearinghouses that transmit health information for purposes related to the provision of healthcare or payment for healthcare services.⁸⁶ The definition of individually identifiable health information includes future health outcomes, since it is defined as

information, including demographic data, that relates to the individual’s past, present or future physical or mental health or condition, the provision of health care to the individual, or the past, present, or future payment for the provision of health care to the individual, and that identifies the individual or for which there is a reasonable basis to believe it can be used to identify the individual.⁸⁷

Under the Privacy Rule, covered entities are required to obtain patients’ permission to disclose their PHI to third parties, and to allow patients to view and obtain copies of their health records, receive accounting of disclosures of their PHI, and request providers to correct errors in their medical records, and to request that providers restrict how their health data are used. Covered entities that suffer data breaches, such as hacking or unintended disclosure, must notify affected individuals and the HHS.⁸⁸ The Security Rule is part of the larger Privacy Rule and promotes secure storage and processing of electronic health information, and delineates administrative, physical, and technical safeguards to protect the confidentiality, integrity, and availability of electronic PHI.⁸⁹

Although HIPAA and its amendments provide valuable protections to

83. The Privacy Rule was later modified in August of 2002. *Id.*

84. *Id.* These provisions will be referred to together as HIPAA.

85. Roger Hsieh, *Improving HIPAA Enforcement and Protecting Patient Privacy in a Digital Healthcare Environment*, 46 LOY. U. CHI. L.J. 175, 176 (2014).

86. 45 C.F.R. § 160.103 (2020).

87. U.S. Dep’t of Health & Human Servs., *Summary of the HIPAA Privacy Rule*, HHS.GOV, <https://www.hhs.gov/hipaa/for-professionals/privacy/laws-regulations/index.html> [<https://perma.cc/L9AM-WKTL>].

88. See 45 C.F.R. §§ 164.508–.510 (2020); Hoffman, *Genetic Testing & PHA Regulation*, *supra* note 48, at 144–45.

89. Hoffman & Podgurski, *supra* note 8, at 336.

individuals disclosing health information, it is limited in scope and has been criticized on multiple grounds. First, HIPAA only applies to covered entities and their business associates within the scope of their healthcare services. HIPAA does not apply to direct-to-consumer health record providers or to employer-provided personal health records that are not connected to an employer health plan.⁹⁰ These include the majority of mobile health and wellness apps and wearables,⁹¹ as well as other parties that possess and handle health information, such as data brokers and marketers. HIPAA does not prohibit these noncovered entities from disclosing health information about patients nor selling information about patients.⁹² Thus, predictive health analytics outcomes generated by these entities will not be protected under current HIPAA, and such entities are free to disclose, publicize, and even sell individually identifiable predictive results.⁹³

Second, HIPAA lacks a private cause of action. Victims cannot recover damages for HIPAA violations, which are subject only to penalties paid to the federal government.⁹⁴ In the absence of a private right of action, HIPAA leaves victims without recovery and inadequately incentivizes covered entities to comply, as evidenced by continuing HIPAA violations.⁹⁵ Third, the Privacy Rule does not enable individuals to establish the origin of their PHI or verify how their information has been used.⁹⁶ Thus, individuals may never discover that their information has been used for improper purposes. Fourth, HIPAA only protects PHI, which by definition only encompasses data that includes one or more of eighteen identifiers.⁹⁷ If these identifiers

90. Hiller, *supra* note 7, at 285.

91. *Id.*; Hoffman, *Genetic Testing & PHA Regulation*, *supra* note 48, at 128.

92. Hoffman, *Genetic Testing & PHA Regulation*, *supra* note 48, at 149.

93. *Id.* Such entities are not covered by HIPAA but are subject to the terms of their user agreements.

94. Morgan Leigh Tendam, Note, *The HIPAA-Pota-Mess: How HIPAA's Weak Enforcement Standards Have Led States to Create Confusing Medical Privacy Remedies*, 79 OHIO ST. L.J. 411, 413 (2018). The regulations leave enforcement in the hands of the Department of Health and Human Services Office for Civil Rights (OCR) and state attorneys general offices, which may or may not have adequate resources for prosecution of HIPAA violations. Hoffman, *Genetic Testing & PHA Regulation*, *supra* note 48, at 145.

95. Tendam, *supra* note 94, at 413.

96. Hoffman & Podgurski, *supra* note 8, at 337.

97. The eighteen identifiers are:

(1) Names (Full or last name and initial)[;] (2) All geographical identifiers smaller than a state, except for the initial three digits of a zip code if, according to the current publicly available data from the U.S. Bureau of the Census: the geographic unit formed by combining all zip codes with the same three initial digits contains more than 20,000 people; and the initial three digits of a zip code for all such geographic units containing 20,000 or fewer people is changed to 000[;] (3) Dates (other than year) directly related to an individual[;] (4) Phone numbers[;] (5) Fax numbers[;] (6) Email addresses[;] (7) Social Security numbers[;] (8) Medical record numbers[;] (9) Health insurance beneficiary numbers[;] (10) Account numbers[;]

are removed from the data, it is not considered PHI and is not subject to the restrictions of the Privacy Rule.⁹⁸ This is a significant risk given that de-identification does not preclude reidentification by data miners and other experts. Research has established that supposedly anonymized data is now commonly identifiable by combining two or more datasets that contain information about the same person, revealing directly identifying information about the individual.⁹⁹ Finally, the Privacy and Security Rules contain many exceptions, such as the ability of covered entities to disclose PHI for the purposes of treatment, payment, and healthcare operations without prior authorization by the individual.¹⁰⁰

2. Health Insurance: The Patient Protection and Affordable Care Act (ACA)

The ACA established guaranteed issuance of health insurance for all individuals who request it.¹⁰¹ The ACA applies to private and federally funded insurance programs. The law has three primary goals: (1) to make affordable health insurance available to more people; (2) to expand the Medicaid program to cover all adults with income below 138% of the federal poverty level; and (3) to support innovative medical care delivery methods designed to lower the costs of healthcare generally.¹⁰² Among other things, the ACA established a new Patient's Bill of Rights and required insurers to make comprehensive insurance coverage accessible to patients.¹⁰³

Four key provisions of the ACA are relevant to future health outcomes: three cover insurance market reforms, and the fourth is the "nondiscrimination" provision in section 1557. Section 1201 of the ACA

(11) Certificate/license numbers[;] (12) Vehicle identifiers (including serial numbers and license plate numbers)[;] (13) Device identifiers and serial numbers[;] (14) Web Uniform Resource Locators (URLs)[;] (15) Internet Protocol (IP) address numbers[;] (16) Biometric identifiers, including finger, retinal and voice prints[;] (17) Full face photographic images and any comparable images[; and] (18) Any other unique identifying number, characteristic, or code except the unique code assigned by the investigator to code the data.

Steve Alder, *What Is Considered Protected Health Information Under HIPAA?*, HIPAA J. (Apr. 2, 2018), <https://www.hipaajournal.com/what-is-considered-protected-health-information-under-hipaa> [<https://perma.cc/A32Y-HDE2>]. PHI also includes genetic information. See 45 C.F.R. §§ 160.103, 164.501 (2020).

98. Alder, *supra* note 97.

99. Boris Lubarsky, *Re-identification of "Anonymized Data,"* 1 GEO. L. TECH. REV. 202, 203 (2017).

100. Hoffman, *Genetic Testing & PHA Regulation*, *supra* note 48, at 145–46.

101. ACA, Pub. L. No. 111-148, sec. 1201, § 2702, 124 Stat. 119, 156 (2008) (codified at 42 U.S.C. §300gg-1 (2018)).

102. U.S. Ctrs. for Medicare & Medicaid Servs., *Affordable Care Act (ACA)*, HEALTHCARE.GOV, <https://www.healthcare.gov/glossary/affordable-care-act> [<https://perma.cc/QD2H-A3W9>].

103. See generally ACA, Pub. L. No. 111-148, 124 Stat. 119 (2010) (codified as amended in scattered sections of 26 and 42 U.S.C.).

amends sections 2701–2708 of the Public Health Service Act (“PHS”).¹⁰⁴ The ACA’s amendment to section 2704 of the PHS prohibits insurers and health insurance issuers from excluding or otherwise discriminating against individuals based on pre-existing health conditions or health status.¹⁰⁵ Second, the amendment to section 2701 of the PHS prohibits insurers from varying premium rates on the basis of health status and allows health insurance issuers to vary premiums only with respect to four variables: (1) whether the plan or coverage covers an individual or family; (2) geographic area; (3) age; and (4) tobacco use.¹⁰⁶ The amendment to section 2701(a)(1)(B) expressly prohibits variation to the particular plan or coverage by any factor not listed.¹⁰⁷ Third, the amendment to section 2705 prohibits discrimination by insurers against individual participants and beneficiaries based on the following factors: (1) health status; (2) medical condition; (3) claims experience; (4) receipt of health care; (5) medical history; (6) genetic information; (7) evidence of insurability; (8) disability; and (9) any other health status related factor determined appropriate by the government.¹⁰⁸ Finally, section 1557 of the ACA prohibits discrimination in the healthcare system on the basis of race, color, national origin, sex, age,¹⁰⁹ or disability.¹¹⁰ Section 1557 is broad and applies to: health providers, programs or activities that receive funding from the federal government, such as hospitals that accept Medicare patients and doctors who treat Medicaid patients; any health program administered by the federal government; and insurers that participate in insurance exchanges (online marketplaces where consumers can compare and buy individual health

104. See ACA sec. 1201, 124 Stat. at 154–61.

105. ACA sec. 1201, § 2704, 124 Stat. at 154–55 (codified at 42 U.S.C. § 300gg-1, gg-3 to gg-4 (2018)).

106. ACA sec. 1201, § 2701(a)(1)(A), 124 Stat. at 155 (codified at 42 U.S.C. § 300gg (2018)). See also U.S. Ctrs. For Medicare & Medicaid Servs., *How Insurance Companies Set Health Premiums*, HEALTHCARE.GOV, <https://www.healthcare.gov/how-plans-set-your-premiums> [<https://perma.cc/Z4TU-R2PU>].

107. ACA sec. 1201, § 2701(a)(1)(B), 124 Stat. at 155 (codified at 42 U.S.C. § 300gg (2018)).

108. ACA sec. 1201, § 2705(a), 124 Stat. at 156 (codified at 42 U.S.C. § 300gg-4 (2018)).

109. Although insurers may vary premiums based on age, they are prohibited from excluding or adversely treating an individual on any of the listed prohibited bases. U.S. Dep’t of Health & Human Servs., *Frequently Asked Questions: Section 1557 of the Affordable Care Act*, HHS.GOV, <https://www.hhs.gov/civil-rights/for-individuals/section-1557/section-1557-proposed-rule-faqs/index.html> [<https://perma.cc/3QSL-C2EN>].

110. Wyatt Foret, *Trans/Forming Healthcare Law: Litigating Antidiscrimination Under the Affordable Care Act*, 28 YALE J.L. & FEMINISM 243, 247 (2017). Section 1557 does not itself enumerate protected categories, but prohibits discrimination on the ground prohibited “under title VI of the Civil Rights Act of 1964 (42 U.S.C. 2000d et seq.), title IX of the Education Amendments of 1972 (20 U.S.C. 1682 et seq.), the Age Discrimination Act of 1975 (42 U.S.C. 6101 et seq.), or section 504 of the Rehabilitation Act of 1973 (29 U.S.C. 794).” *Id.*

insurance plans).¹¹¹

Despite these protections, consumers still face obstacles in obtaining adequate and affordable health coverage, and the ACA provides those injured by insurer violations with limited options for recourse.¹¹² One major criticism of the ACA is that, like HIPAA, it lacks a private right of action for individuals harmed by insurer violations under the private insurance market reform provisions of the Act.¹¹³ The only exception is section 1557, which allows private claims against facially neutral healthcare policies, treatment decisions, and insurance coverage on the basis that they disproportionately affect members of protected classes.¹¹⁴ Consumers may appeal certain denials of benefits using their health plans' internal procedures and, only if that fails, ask an external reviewer to assess their claims.¹¹⁵ If the appeal is unsuccessful, consumers may lodge a complaint with state and federal regulators, but consumers cannot currently sue under federal law. Thus, consumers can neither compel insurers to comply with the ACA's insurance reform provisions nor collect damages from them if they are harmed by noncompliance.¹¹⁶

3. Nondiscrimination Laws: The ADA and GINA

i. The Americans with Disabilities Act

The Americans with Disabilities Act ("ADA") prohibits employers and other public and private entities from discriminating against individuals because of their physical or mental disabilities. The ADA was enacted in 1990 for the purpose of ensuring that people with physical or mental disabilities are afforded the same rights and opportunities as everyone else.¹¹⁷ Title I of the ADA prohibits discrimination on the basis of disability by employers with fifteen or more employees, and all employment agencies,

111. *Id.*; U.S. Dep't of Health & Human Servs., *Section 1557 of the Patient Protection and Affordable Care Act*, HHS.GOV, <https://www.hhs.gov/civil-rights/for-individuals/section-1557/index.html> [<https://perma.cc/8X3T-KKVQ>].

112. Christine H. Monahan, *Private Enforcement of the Affordable Care Act: Toward an "Implied Warranty of Legality" in Health Insurance*, 126 YALE L.J. 1118, 1122 (2017).

113. *Id.* at 1123 n.23 (noting that Section 1557 has been interpreted to create a private right of action, at least with respect to claims of disparate impact discrimination). *See also* ACA, Pub. L. No. 111-148, sec. 1201, §§ 2701, 2704, 2705, 124 Stat. 119, 154-60 (2008) (codified at 42 U.S.C. 300gg to gg-4).

114. Monahan, *supra* note 112, at 1123.

115. *Id.* at 1122-23.

116. *Id.*

117. Mid-Atl. ADA Ctr., *An Overview of the Americans with Disabilities Act*, ADA NAT'L NETWORK, <https://adata.org/factsheet/ADA-overview> [<https://perma.cc/A92Y-QYNS>].

labor organizations, and joint-labor management companies.¹¹⁸

Title I of the Act is relevant to discrimination on the basis of predicted health outcomes, since employers are likely to have particular interest in individuals' future health. The ADA defines disability as "(A) a physical or mental impairment that substantially limits one or more major life activities of such individual; (B) a record of such an impairment; or (C) being regarded as having such an impairment."¹¹⁹ For the purposes of (C), an individual is regarded as having an impairment if he or she establishes discrimination based on an actual or *perceived* physical or mental impairment, regardless of whether the impairment actually limits or is perceived to limit a major life activity, excluding "transitory and minor" impairments (those with an actual or expected duration of six months or less).¹²⁰ Thus, under the "regarded as" provision, workers who have impairments that are not severe enough to be considered disabilities or those who do not have disabling conditions but are wrongly perceived as being impaired are also protected under the law.¹²¹ Although this definition is broad in scope, the ADA does not expressly prohibit discrimination against individuals who have never had disabilities but are at an increased risk of developing a disability later in life. Thus, the ADA does not bar employers and other covered stakeholders from discriminating against employees based on predicted health outcomes.¹²² For example, employers may be permitted to base employment decisions on their concern that candidates or employees may become ill in the future based on predictive analytics models, which take into account a variety of factors such as diet, social media use, and purchasing patterns.¹²³

Two major limitations of the ADA are that (1) it is a voluntary compliance law, and (2) the ADA only provides for injunctive relief and attorney's fees to plaintiffs who successfully sue and win their cases.¹²⁴

118. Americans with Disabilities Act, 42 U.S.C. §§ 12111–12112, 12132 (2018).

119. 42 U.S.C. § 12102.

120. 42 U.S.C. § 12102(3). "[M]ajor life activities include, but are not limited to, caring for oneself, performing manual tasks, seeing, hearing, eating, sleeping, walking, standing, lifting, bending, speaking, breathing, learning, reading, concentrating, thinking, communicating, and working." § 12102(2)(A). Major bodily functions include "but [are] not limited to, functions of the immune system, normal cell growth, digestive, bowel, bladder, neurological, brain, respiratory, circulatory, endocrine, and reproductive functions." § 12102(2)(B).

121. Hoffman, *Big Data and the ADA*, *supra* note 76, at 786.

122. Sharona Hoffman, *Big Data's New Discrimination Threats: Amending the Americans with Disabilities Act to Cover Discrimination Based on Data-Driven Predictions of Future Disease*, in *BIG DATA, HEALTH LAW, AND BIOETHICS* 85, 92 (Glenn Cohen, Allison Hoffman & William Sage eds., 2018).

123. *Id.*

124. *Id.*

Thus, unlike with other civil rights laws, employers are not required to report on their compliance with the ADA, and violations are only discovered if employees are wrongfully treated and choose to sue their employer. Moreover, many employees, particularly those who are discharged, do not have the time or resources to bring suit, and employers may not think it is worth complying with the ADA because the likelihood of being found liable is so small.¹²⁵

ii. The Genetic Information Nondiscrimination Act

The Genetic Information Nondiscrimination Act of 2008 (“GINA”) protects individuals against discrimination by insurance companies and employers on the basis of genetic information.¹²⁶ Such genetic discrimination is defined as the “adverse treatment of an individual based on genotype”¹²⁷ or the “use of genetic information to draw a distinction among individuals or groups, plus an element of irrationality, social unacceptability, or both.”¹²⁸

Title I of GINA prohibits discrimination in health insurance on the basis of genetic information,¹²⁹ just as the ACA subsequently extended this prohibition to discrimination based on health status generally. Prior to GINA and the ACA, health insurers could request genetic testing and engage in risk rating based on such information.¹³⁰ Although HIPAA prohibits the use of genetic information in decisions about eligibility or premiums, it does not apply to the individual health insurance market or prohibit insurers from requesting or requiring genetic information.¹³¹ Under Title I, insurers are barred from (1) using genetic information to determine coverage, eligibility, or premiums; (2) requesting or requiring genetic testing or genetic information; and (3) obtaining genetic information for underwriting purposes.¹³² Thus, GINA prevents insurers from applying pre-existing

125. Julian Cardillo, *The Americans with Disabilities Act: 25 Years Later*, BRANDEIS: NOW (July 23, 2015), <https://www.brandeis.edu/now/2015/july/parish-ada-qanda.html> [<https://perma.cc/6YP4-3V9Y>].

126. Genetic Information Nondiscrimination Act (GINA) of 2008, Pub. L. No. 110-123, § 101, 122 Stat. 881, 883–88 (codified as amended in scattered sections of 29 U.S.C.).

127. Mark A. Rothstein, *GINA's Beauty Is Only Skin Deep*, GENEWATCH, Apr.–May 2009, at 9, 9.

128. Mark A. Rothstein, *GINA at Ten and the Future of Genetic Nondiscrimination Law*, HASTINGS CTR. REP., May–June 2018, at 5, 5 [hereinafter, Rothstein, *GINA at Ten*].

129. *Id.*

130. Bradley A. Areheart & Jessica L. Roberts, *GINA, Big Data, and the Future of Employee Privacy*, 128 *Yale L.J.* 710, 722 (2019).

131. *Id.*

132. GINA, Pub. L. No. 110-123, §§ 101–106, 122 Stat. 881, 883–905 (2008) (codified as amended in scattered sections of 26, 29, and 42 U.S.C.); see also Hoffman, *Genetic Testing & PHA Regulation*, *supra* note 48, at 143.

condition exclusions for coverage, requesting genetic information, and using that information in their coverage decisions. However, GINA does not enable independent enforcement mechanisms for its health insurance provisions since it relies on enforcement mechanisms of those underlying laws (specifically HIPAA and the ACA), most of which do not have a private right of action.¹³³

Title II of GINA prohibits employers with fifteen or more employees, employment agencies, labor organizations, joint labor-management training programs, and apprenticeship programs from discriminating against candidates and employees on the basis of genetic information.¹³⁴ Title II makes it unlawful for an employer to fail or refuse to hire, discharge, or otherwise discriminate against any employee with respect to compensation, terms, conditions, or privileges of employment due to genetic information.¹³⁵ Employers are also prohibited from limiting, segregating, or classifying employees in any way that would deprive or tend to deprive them of opportunities or otherwise adversely affect their status because of genetic information.¹³⁶ Title II further prohibits employers from requesting, requiring, or purchasing genetic information, although this prohibition is subject to numerous exceptions.¹³⁷ Unlike the insurer provisions in Title I, Title II's employment provisions enable an independent, private right of action by employees against employers allegedly in breach of Title II.¹³⁸ However, unlike other antidiscrimination laws, GINA does not allow discrimination to be proven based on disparate impact and prohibits only explicit discrimination on the basis of genetic information.¹³⁹ An individual who believes that an employer has violated GINA may file a charge of discrimination with the U.S. Equal Employment Opportunity Commission ("EEOC"), which will then attempt to settle or resolve the dispute through mediation with the employer.¹⁴⁰ If the dispute is not resolved, the EEOC will conduct further investigation and either notify the person of his or her right

133. Areheart & Roberts, *supra* note 130, at 716.

134. GINA §§ 201–202, 42 U.S.C. §§ 2000ff to 2000ff-1 (2018).

135. § 202(a)(1).

136. *Id.*

137. *Id.*

138. *See* GINA § 207, 42 U.S.C. § 2000ff-6 (2018).

139. Jessica L. Roberts, *Preempting Discrimination: Lessons from the Genetic Information Nondiscrimination Act*, 63 VAND. L. REV. 439, 452–53 (2010) [hereinafter Roberts, *Preempting Discrimination*].

140. U.S. EQUAL EMP'T OPPORTUNITY COMM'N, WHAT YOU SHOULD KNOW: QUESTIONS AND ANSWERS ABOUT THE GENETIC INFORMATION NONDISCRIMINATION ACT (GINA) AND EMPLOYMENT (2013), <https://www.eeoc.gov/laws/guidance/what-you-should-know-questions-and-answers-about-genetic-information> [<https://perma.cc/6KLR-AYK9>].

to file a lawsuit or file suit on the person's behalf.¹⁴¹

GINA has been criticized on several grounds. First, as discussed, Title I does not enable a private right of action.¹⁴² GINA's two titles are also separated by a "firewall," in that claimants cannot sue under both provisions simultaneously to reap the benefits of the two sections' separate remedies.¹⁴³ For example, employment discrimination claims under Title II are prohibited for conduct that is actionable under the health insurance provisions in Title I unless the employer independently violated Title II.¹⁴⁴ Second, GINA is limited to health insurance and employment and does not cover many other forms of discrimination, such as in life insurance, disability insurance, long-term care insurance, mortgage insurance, educational opportunities, and commercial and real property transactions.¹⁴⁵ Thus, it is perfectly legal for banks, landlords, schools, and life insurers to discriminate on the basis of genetic information, should they obtain it. Third, GINA takes a narrow view of "genetic" information. "Section 101(d)(7)(A) defines a 'genetic test' as 'an analysis of human DNA, RNA, chromosomes, proteins or metabolites that detects genotypes, mutations, or chromosomal changes.'" ¹⁴⁶ This definition appears to leave out epigenetic marks, microbiome data, and various other biological measures that have emerged since GINA's enactment in 2008.¹⁴⁷ Moreover, as opposed to the ADA, which applies only to substantial impairments that have manifested or are regarded as having manifested, GINA also only applies to asymptomatic individuals.¹⁴⁸ Although the ADA and GINA appear to be mirror images of each other regarding the stage of disease or disability progression, there is a substantial gap in coverage between the two antidiscrimination laws, such as coverage for individuals who have a biomarker or subclinical marker of gene expression.¹⁴⁹ Finally, GINA's employment provision contains numerous exceptions permitting an employer to obtain genetic information under certain conditions, including inadvertent disclosure, through voluntary wellness programs, when processing medical leave, and through occupational monitoring of toxic substances.¹⁵⁰

141. *Id.*

142. Areheart & Roberts, *supra* note 130, at 716.

143. Roberts, *Preempting Discrimination*, *supra* note 139, at 453.

144. *Id.* at 450.

145. Rothstein, *GINA at Ten*, *supra* note 128, at 5.

146. *Id.* at 5–6.

147. *Id.* at 6.

148. *Id.*

149. *Id.*

150. Areheart & Roberts, *supra* note 130, at 728.

B. GAP ANALYSIS

The aforementioned laws provide a patchwork of protection over health information against breaches in privacy and discriminatory use by health insurers, providers, and employers. What gaps remain in current legislation over the protection of predictive health information?

1. HIPAA

Since HIPAA's safeguards for the privacy and security of PHI encompasses information relating to the "past, present, or future" health or condition of an individual, predicted health outcomes would appear to be covered as they relate to an individual's future health or condition.¹⁵¹ However, predicted health outcomes under HIPAA are subject to the statutes' limited nature: it only applies to covered entities (health plans, healthcare clearinghouses, and healthcare providers), and their business associates who store or transmit health information related to the provision of healthcare or payment for healthcare services.¹⁵² Thus, HIPAA leaves without protection most information that can be used to predict an individual's future health and that is possessed by employers, lenders, life insurance companies, and data brokers, as well as collected by the majority of mobile health apps. Moreover, were a person to discover that information regarding his or her predicted health had been impermissibly disclosed under HIPAA, the person would not be able to sue for damages under federal law, regardless of the consequences suffered as a result of the disclosure. Given the sensitivity of an individual's predicted health, particularly if those predictions are of serious or terminal illness, the remedy does not seem to match the level of risk. Thus, HIPAA falls short of providing individuals with comprehensive protection over their health information, including predicted health outcomes.

2. The ACA

The ACA's limited scope is similar to HIPAA's. The ACA only protects against differential treatment by health insurers in determining issues such as coverage and premium setting. Additionally, the ACA does not expressly prohibit health insurers from excluding or otherwise discriminating against an individual on the basis of predictive health information. The express language of the statute prohibits insurers from denying people coverage based on pre-existing conditions or health status,

151. 45 C.F.R. § 160.103 (2020).

152. *Id.*

that is, health problems that an individual experienced before the date that their new health coverage is to start. However, health insurers are limited to four demographic variables (individual or family, geography, age, and tobacco use) in setting and determining premiums. The question, then, is whether the prohibition on using any information related to an individual's "health status" encompasses information generated by an algorithm regarding the health outcomes predicted for the individual. The ACA does not expressly allow insurers to use this information in determining coverage or premiums, but it does not forbid insurers from doing so.

3. The ADA

The ADA, which applies to employers, public services, and places of accommodation, does not ban discrimination against individuals based on their risk of future disease. Accordingly, under the plain language of the ADA, employers are free to base their employment decisions on predicted health outcomes and disqualify applicants or terminate employees due to concern that an employee or potential employee is at a higher risk of developing a disease in the future. However, it is possible that the ADA's "regarded as" provision could include individuals who are perceived as likely to develop physical or mental impairments in the future. Under that provision, a claimant need only prove that the employer discriminated against them on the basis of a nontransitory physical or mental impairment, regardless of whether the employee was healthy in reality.

4. GINA

GINA is the only nondiscrimination statute that was designed specifically to apply to future health conditions. GINA does not, however, provide protection against discrimination based on predictive analytical information because it only covers genetic information. "The statute specifically excludes any information about sex, age, or an individual's own health conditions from its definition of genetic information."¹⁵³

5. Nonlegal Protections

Nonlegal protections, such as market forces and industry self-regulation, may also protect individuals against discrimination based on future health status. For example, health care professionals may adopt their own safeguards in order to minimize the risks of predicted health outcomes

153. Areheart & Roberts, *supra* note 130, at 717.

for patients¹⁵⁴ and to ward off the need for federal or state legislation and regulation. Professional organizations may also develop practice guidelines recommending when clinicians should or should not obtain certain predictions. Corporations may also engage in self-governance. For example, health insurers and employers, particularly those that are public corporations, are required by the Securities Exchange Commission (“SEC”) to disclose the fundamental values by which they operate. Technology market leaders have recognized the need for principled governance over predictive analytics. For example, in March 2019, Google announced the launch of an “Advanced Technology External Advisory Council” that would provide ethical oversight of the technology giant’s development of artificial intelligence technologies.¹⁵⁵ Although this ethics board was canceled before its creation,¹⁵⁶ the company has created exemplary guidelines around its ethical objectives regarding artificial intelligence.¹⁵⁷

C. SUMMARY

HIPAA is the only statute that covers, by its express language, an individual’s future health status based on nongenetic risk factors. However, HIPAA’s coverage is limited, and it only provides for violations of privacy, rather than discriminatory treatment. The ACA does not expressly prohibit covered insurers from discriminating against individuals based on their predicted health outcomes, although it does not allow them to consider this information when determining coverage and premiums. Similarly, the ADA does not address future health status, although the regarded as provision may be interpreted to cover and individuals’ future health. Finally, GINA only covers genetic information, leaving the majority of data sources of predictive analytics models unprotected. GINA is thus of interest principally because of the similarities between genetic testing and predictive analytics in forecasting the risk of future disease. The next Section of this Note will explore these similarities in detail.

III. PART III

In this Part, I will discuss similarities between genetic testing and predictive analytics and provide recommendations for future regulation of predictive health analytics.

154. See Hoffman, *Genetic Testing & PHA Regulation*, *supra* note 48, at 161.

155. *Id.* at 162.

156. *Id.*

157. *Our Principles*, GOOGLE AI, <https://ai.google/principles> [<https://perma.cc/QN84-CJJN>].

A. SIMILARITIES BETWEEN GENETIC TESTING & PREDICTIVE ANALYTICS

First, predictive analytics and genetic testing are used to predict an individual's risk of developing a disease. Genetic tests can tell people whether they have an increased proclivity for disease with hereditary linkages based on certain genetic mutations, such as the BRCA 1 or 2 genes and breast cancer.¹⁵⁸ In this sense, genetic testing is a type of predictive analytics to forecast the likelihood of future disease based on DNA because the identification of relevant genes often depends on large databases from which all linkages can be identified. Indeed, some diseases, such as Alzheimer's and heart disease, may be predicted based on information produced either through genetic analysis or predictive analytics, with the former using DNA variation and the latter using other types of health information. Like predictive analytics, many genetic tests cannot tell a person with certainty whether he or she will develop a disease, but rather indicate that the person is at a significantly increased risk of the disease. Many genetic mutations are not completely penetrant, so that whether an individual with the mutation will develop the condition will depend on other genetic and environmental factors. Conversely, a genetic test that reveals that an individual does not have a gene that increases the risk of a particular disease does not rule out the individual from developing that disease.¹⁵⁹

Second, as a result of this shared predictive power, genetic testing and predictive analytics have raised similar concerns regarding discrimination. Prior to the enactment of GINA, many policymakers and legislators feared that without comprehensive privacy protection, sensitive genetic information could be used for discriminatory purposes.¹⁶⁰ For example, Americans were concerned in a pre-ACA era that a health insurance company that obtained information from a genetic test that revealed an applicant's future disease risk could potentially decline to insure the individual, increase premiums, or otherwise to discriminate against the person.¹⁶¹ Employees and potential job candidates also worried that employers could also obtain genetic information through medical examinations or wellness programs and likewise choose to fire them, reject potential candidates, or otherwise segregate groups within

158. Areheart & Roberts, *supra* note 130, at 714.

159. Hoffman, *Genetic Testing & PHA Regulation*, *supra* note 48, at 137–38.

160. *See id.* at 139. Survey data from 2004 also indicated that 68% of respondents opposed allowing health insurers access to their genetic information, and 85% opposed allowing employers access to genetic information. GENETICS & PUB. POLICY CTR., PUBLIC AWARENESS AND ATTITUDES ABOUT REPRODUCTIVE GENETIC TECHNOLOGY 12–13 (2002), <https://jscholarship.library.jhu.edu/bitstream/handle/1774.2/979/PublicAwarenessAndAttitudes.pdf?sequence=1&isAllowed=y> [<https://perma.cc/4GLS-4DED>].

161. Hoffman, *Genetic Testing & PHA Regulation*, *supra* note 48, at 140.

the workplace based on their genetic information.¹⁶²

GINA was the first major antidiscrimination statute of the new millennium, and unlike other civil rights laws, such as the ADA, GINA was not enacted to redress a history of discrimination in employment, in insurance, or otherwise.¹⁶³ Rather, Congress believed that individuals would be more willing to avail themselves of genetic testing and related services if the law contained protections against misuse of genetic information.¹⁶⁴ Policymakers and legislators were convinced that as EHRs and other health records came to routinely contain genetic information and genetic testing was performed on a widespread basis, individuals could face discrimination based on their genetic information.¹⁶⁵ Proponents of genetic testing urged that legal protections were needed to assure patients and other members of the public that they can safely undergo genetic tests in clinical care, participate in genetic research, and utilize genetic based treatments without fear that their genetic information will be used against them.¹⁶⁶ Similarly, individuals should be encouraged to engage with the healthcare system (which is increasingly relying on a “learning health systems” model that uses data to improve quality, cost, and so forth) and to allow their health data to be used in predictive analytics research and programs without having to worry that information regarding their possible future health outcomes will be misused by health insurers, providers, and employers. This is particularly important since predictive analytics is powered by data, which needs to be collected from and verified on patients in order to continue to advance and improve models.¹⁶⁷

Third, genetic testing and predictive health analytics pose similar issues around disclosure (or nondisclosure) of predicted health outcomes to patients. Studies on the psychological effects of genetic testing for the BRCA genes have shown that individuals who discover they are at an increased risk for breast cancer experience distress, depression, guilt, and lowered self-esteem, goals, and expectations.¹⁶⁸ Incidental findings are also increasingly

162. *Id.* at 139–40.

163. See generally Mark A. Rothstein, Jessica Roberts & Tee L. Guidotti, *Limiting Occupational Medical Evaluations Under the Americans with Disabilities Act and the Genetic Information Nondiscrimination Act*, 41 AM. J.L. & MED. 523 (2015).

164. *Id.* at 526.

165. *Id.*

166. Rothstein, *GINA at Ten*, *supra* note 128; Genetic Information Nondiscrimination Act of 2008, Pub. L. No. 110-123, § 2(5), 122 Stat. 881, 882–83 (explaining that GINA was enacted in order to allow “individuals to take advantage of genetic testing, technologies, research, and new therapies.”).

167. See WATSON, *supra* note 66, at 14.

168. Katherine A. Schneider, *Adverse Impact of Predisposition Testing on Major Life Activities*:

prevalent as genetic tests have shifted from discrete tests to large-scale genetic sequencing (or “whole genome sequencing”), which has the potential to result in large numbers of incidental findings. While some genetic variants discovered reveal clinically relevant information, much is of unknown or uncertain medical value.¹⁶⁹ Moreover, as science evolves, variants that are of unknown significance today may later be found to be associated with critical diseases or conditions.¹⁷⁰ As opposed to other types of diagnostic tests, such as MRIs and CT scans, genome sequencing can lead to dozens or hundreds of disparate incidental findings due to the large number of base pairs analyzed.¹⁷¹ Similarly, due to the vast amounts of data used to power predictive analytics models, the models may potentially result in a similarly large amount of incidental findings.

B. DIFFERENCES BETWEEN GENETIC TESTING & PREDICTIVE ANALYTICS

Although genetic testing and predictive analytics share many similarities due to their shared predictive power, the two also differ in some key respects. Based on these differences, should they be afforded different protections under the law?

1. Diagnostic Screening Versus System-Enhancing Tool

One critical difference between genetic testing and predictive health analytics is that genetic testing is a diagnostic screening conducted to determine an individual’s chance of developing a genetic disorder, whereas predictive analytics models are developed by the healthcare system to achieve improved outcomes and efficiency for patient populations overall. Moreover, genetic testing, like most other diagnostic tests, requires consent on behalf of the patient.¹⁷² Informed consent for genetic testing is generally obtained during an office visit by a doctor or genetic counselor, who

Lessons from BRCA1/2 Testing, 3 J. HEALTH CARE L. & POL’Y 365, 369, 372–74 (2000).

169. Disclosure of incidental findings from genetic testing has sparked discussion around an individual’s “right not to know.” The right not to know has been argued as an enhancement of autonomy, since the decision to know or not to know is taken out of the hands of the doctor by the patient. Scholars argue that one’s right to self-determination implies a right to make decisions about learning, or not learning, sensitive medical information. R. Andorno, *The Right Not to Know: An Autonomy Based Approach*, 30 J. MED. ETHICS 435, 435 (2003). The issue of whether a patient has a right to know or not to know of predicted health outcomes is likely to generate similar issues, especially as the predictive power and range of results generated by predictive analytics increases.

170. See INCIDENTAL FINDINGS, *supra* note 65, at 2.

171. *Id.* at 35.

172. U.S. Nat’l Library of Med., *What is Informed Consent?*, MEDLINE PLUS, <https://ghr.nlm.nih.gov/primer/testing/informedconsent> [<https://perma.cc/Y65L-FY84>].

discusses the test and the possible consequences of test results.¹⁷³ On the other hand, developers of predictive analytics models are generally permitted to use patient data that have already been collected without explicit consent.¹⁷⁴ The informed consent of human subjects is not required if researchers cannot identify the individuals whose data are being analyzed.¹⁷⁵

For many predictive analytics models, obtaining informed consent from individuals whose data is being used to develop the models is impracticable due to the sheer volume of individuals from which data are acquired.¹⁷⁶ For example, a recent study using a machine learning algorithm to identify undiagnosed Alzheimer's patients used data on nearly 90 million subjects in the United States.¹⁷⁷ The researchers concluded that applying the model to U.S. residents could drive major advances in Alzheimer's research by enabling more accurate and earlier diagnosis, facilitate timely referral to expert sites, and increase potential enrollment in clinical trials.¹⁷⁸ The research did not address, however, whether individuals would need to consent to their data being processed by the algorithm, whether they would have a choice in receiving (or not receiving) a diagnosis, or whether use of the algorithm would be disclosed to patients.¹⁷⁹

2. Impact on Relatives

Genetic information also differs from other types of health information due to its implications for biologically related family members of a person tested.¹⁸⁰ While, on one hand, DNA is a unique identifier of each person, genetic data are not limited to the individual, and information about an individual's DNA may reveal information about his or her relatives. This "mixed nature" of genetic information implicates disclosure and privacy concerns.¹⁸¹ Clinicians have a duty to disclose the results of genetic tests to their patients, but are not required or permitted to inform the genetically at-

173. *Id.*

174. Cohen et al., *supra* note 1, at 1139, 1141.

175. *See* 45 C.F.R. § 46.104.

176. *See* O. Uspenskaya-Cadoz, C. Alamuri, L. Wang, M. Yang, S. Khinda, Y. Nigmatullina, T. Cao, N. Kayal, M. O'Keefe & C. Rubel, *Machine Learning Algorithm Helps Identify Non-Diagnosed Prodromal Alzheimer's Disease Patients in the General Population*, 6 J. PREV. ALZHEIMER'S DISEASE 185, 185 (2019).

177. *Id.*

178. *Id.*

179. *See id.*

180. *See* INCIDENTAL FINDINGS, *supra* note 65, at 36.

181. Ellen Wright Clayton, Barbara J. Evans, James W. Hazel & Mark A. Rothstein, *The Law of Genetic Privacy: Applications, Implications, and Limitations*, 6 J.L. & BIOSCIENCES 1, 2 (2019).

risk relatives of their patients without prior consent from the patient.¹⁸² However, there has been disagreement in the field as to whether clinicians should have an obligation to alert at-risk family members of particular genetic predispositions, or to recommend that individuals tested inform their at-risk family members of their own results.¹⁸³ Moreover, if an individual's genetic test results are disclosed, such results could also lead to differential treatment of his or her family members based on that information. GINA addresses some of these concerns, as its privacy provisions prohibit employers and health insurers from requesting, requiring, or purchasing genetic information with respect to an employee or covered individual (or potential employee / covered individual) or a *family member* of said person.¹⁸⁴ GINA does not, however, apply to parties other than health insurers, health plans, and employers, such as healthcare providers (doctors, nurses, and so forth), lenders, and life insurance companies.

Predicted health outcomes from predictive analytics models differ from predispositions screened for by genetic tests in that they can be generated by a wide variety of nongenetic data sources. Thus, such outcomes may not have the same implications for biological relatives. However, even though genetic information is afforded special protections under the law, one practical problem with its separate treatment is the difficulty in isolating it from other types of medical information in health records.¹⁸⁵ For example, a patient's own medical history may imply much about their genetic makeup, and family health history information often contains, or is sourced from, the person's genetic information.¹⁸⁶ Such information is widely dispersed in health records and can be used to generate predicted health outcomes. Thus, this difference may not be significant in determining differential protection of predictive analytics results as compared to genetic information.

C. RECOMMENDATIONS FOR FUTURE PREDICTIVE HEALTH ANALYTICS REGULATION

Given significant similarities between genetic testing and predictive health analytics, the laws enacted to protect persons from harm from genetic testing seem to provide a fitting framework for predictive health analytics legislation. This Section will provide recommendations for future predictive health analytics legislation based on lessons from genetic testing and GINA.

182. *Id.* at 15.

183. *Id.*

184. *Id.* at 13.

185. *See id.* at 7–8.

186. *Id.* at 8 n.37.

1. Antidiscrimination Legislation

GINA has been termed a “preemptive” civil rights law in that it was enacted before the data produced through genetic screening and diagnosis had resulted in widespread discrimination in health insurance and employment.¹⁸⁷ Some scholars argue that the time is ripe for the law to similarly “preemptively” intervene in the predictive health analytics field.¹⁸⁸ However, by the time GINA was enacted, genetic discrimination in both health insurance and employment had been reported and litigated in several court cases, which suggests that the “preemptive” description of GINA is an overstatement.¹⁸⁹ Studies on medical underwriting also indicate that a significant percentage of health insurance applicants were denied coverage, administered higher premiums, or given limited benefits based on their genetic information.¹⁹⁰ Francis Collins, then leader of the Human Genome Project and Director of the National Institutes of Health, stated, “Genetic information and genetic technology . . . can be used in ways that are fundamentally unjust. . . . Already . . . people have lost their jobs, lost their health insurance, and lost their economic well-being . . . due to the unfair and inappropriate use of genetic information.”¹⁹¹ Polling data amongst Americans also found that 85% of Americans believed that without amending current law, employers would use this information to discriminate amongst employees.¹⁹²

This does not appear to be the case with predictive analytics today (yet, at least). Although predictive health analytics poses a potential for discrimination in health insurance and the workplace in the future, actual harm has not yet been reported. Survey data of Americans’ thoughts regarding predictive health analytics does not yet publicly exist. The majority of predictive health analytics models remain in pilot stages or in research settings, but policymakers and legislators should be prepared for reactions similar to those in response to genetic testing as predictive models expand into clinical and commercial environments. There is also a risk that subjects of discrimination based on predicted health outcomes may never know that their predicted health outcomes were used against them. Although

187. Roberts, *Preempting Discrimination*, *supra* note 139, at 462.

188. See, e.g., Hoffman, *Genetic Testing & PHA Regulation*, *supra* note 48, at 129.

189. Roberts, *Preempting Discrimination*, *supra* note 139, at 462–63.

190. Louise Slaughter, *Genetic Information Non-Discrimination Act*, 50 HARV. J. ON LEGIS. 41, 46 (2013).

191. *Id.* at 46 (citing *Genetic Information in the Workplace: Hearing Before the S. Comm. on Health, Educ., Labor, and Pensions*, 106th Cong. 7 (2000) (statement of Dr. Francis S. Collins, Dir. of the Nat’l Human Genome Research Inst.)).

192. *Id.*

only deidentified patient level data may be used without a patient's consent, reidentification may occur.¹⁹³ While few individuals have the resources or motivation to engage in reidentification, it is still possible for breaches to occur. Accordingly, bioethicists have recommended that collectors of predictive health analytics data notify patients, where possible, that data gathered on them in the course of their healthcare may be used in deidentified form in predictive analytics models.¹⁹⁴ Healthcare insurers and providers who are developing predictive analytics models are likely able to contact their members, and doing so would be akin to the regular practice of a hospital telling patients that their records may be used for quality improvement.¹⁹⁵

2. Disclosure of Predictive Health Analytics Outcomes

Disclosure of predictive health analytics may also be a first step to addressing potential psychological harms to patients. Disclosure may also help to clarify the corresponding duty of clinicians to inform their patients of future health risks discovered by predictive analytics models. Practitioners must anticipate situations in which they become privy to a patient's future health outcomes without that individual having actively sought out this information. Drawing a parallel to the recommended treatment of incidental findings by the Presidential Commission for the Study of Bioethical Issues, clinicians should engage in open communication with their patients to ensure that individuals are informed of their plan for disclosing and managing predicted health findings, including which findings will and will not be returned to the patient.¹⁹⁶

Health care professionals' duty to disclose the results of genetic tests to patients' relatives may provide a useful framework for a provider's ethical duty to disclose predictive health analytics results. The relationship between (1) a provider and a patients' relative and (2) a provider that is privy to a patient's predictive health analytics results are similar in that, in both circumstances, the provider and individual may not have a pre-existing relationship.¹⁹⁷ Thus, the standard duty to inform the individual may not exist. The American Society of Human Genetics has recommended a two-part approach to disclosure of genetic test results to relatives.¹⁹⁸ First, under

193. Cohen et al., *supra* note 1, at 1141.

194. See, e.g., *id.*

195. *Id.*

196. INCIDENTAL FINDINGS, *supra* note 65, at 44–45.

197. Kristin E. Schleiter, *A Physician's Duty to Warn Third Parties of Hereditary Risk*, 11 *AMA J. ETHICS* 697, 697 (2009).

198. See *id.* at 699.

a standard duty of care, providers must inform patients prior to genetic testing about the impact of genetic testing on their family members.¹⁹⁹ Second, after satisfying its duty to warn the patient, the provider may use their discretion to notify at-risk members of the patient's family when four factors are present: (1) attempts to encourage disclosure by the patient have failed; (2) harm to the relative is highly likely to occur and is serious and foreseeable; (3) the at-risk relative is identifiable; and (4) the disease is preventable, treatable, or medically accepted standards indicate that early monitoring will reduce the risk of harm.²⁰⁰

Providers can take a similar approach to disclosure of predictive health analytics results by first explaining to their patients that they may become privy to information regarding that patient's future health. Patients could then be given the option of opting into disclosure of such results, specifying the type of results they wish to receive, or both. If patients do not wish to receive results, disclosure to patients without their consent would only be warranted at the provider's discretion if, similar to factors (2) and (4) regarding disclosure of genetic test results to relatives, harm to the patient is highly likely to occur, serious and foreseeable, and disclosure of the condition is likely to reduce the risk of harm through treatment or prevention.²⁰¹

3. Post-ACA Considerations

Future legislation should also take into account changes in the law since GINA's enactment. The ACA, passed two years after GINA, made redundant GINA's health insurance provisions by banning exclusions based on pre-existing conditions and other medically relevant information which would encompass genetic conditions and risks.²⁰² The ACA also prohibits insurance companies from using wellness programs to rate the risks of its

199. *Id.*

200. *Id.* Similarly, the President's Commission for the Study of Ethical Problems in Medicine and Biomedical and Behavioral Research recommended that disclosure without a patient's consent is justified only if:

(1) reasonable efforts to elicit voluntary consent to disclosure have failed; (2) there is a high probability both that harm will occur if the information is withheld and that the disclosed information will actually be used to avert harm; (3) the harm the identifiable individuals would suffer would be serious; and (4) appropriate precautions are taken to ensure that only the genetic information needed for diagnosis or treatment of the disease in question is disclosed.

Id.

201. Factors (1) and (3), attempt to encourage the patient to notify their at-risk family members, and identifiability of that relative are not applicable since it is the patient, not their relative, at issue in the case of predictive health analytics discussed above.

202. Areheart & Roberts, *supra* note 130, at 750.

members.²⁰³ Consequently, protections against discrimination in health insurance based on future health outcomes is likely to be unnecessary in light of the ACA. Thus, if enacted, future legislation should target groups not addressed by the ACA, such as employers, who are incentivized to select candidates with high productivity and low absenteeism. The ADA protects against discrimination by employers on the basis of current disability and GINA against genetic discrimination, but the laws do not address differential treatment based on nongenetic, predicted health outcomes.²⁰⁴

4. Life Insurers and Other Stakeholders

Finally, no existing laws address data privacy and discrimination by other stakeholders with an interest in an individual's future health status, such as life insurance companies. Life insurers, as opposed to health insurers, are permitted to deny coverage based on health status or lifestyle factors. Accordingly, life insurance companies may be incentivized to reidentify deidentified patient data sets sold by data brokers in order to determine an individual's risk of future disease, especially since insured individuals are not required to disclose medical conditions that arise after they purchase a policy.²⁰⁵ Addressing such risks may also require expansion of healthcare information privacy laws. For example, HIPAA could be expanded to cover any party who comes into possession of or stores health information.²⁰⁶ This change would not prevent healthcare providers from contracting with business associates to conduct predictive health analytics, nor prevent data brokers from accessing much of the online and public record data they already do.²⁰⁷ However, it would prevent covered entities, which would include data brokers under the expanded definition, from disclosing health predictions to third parties without the data subject's consent (for example, data brokers and life insurance companies). Entities would be required to inform all data subjects of disclosures of their health information and would have to adhere to the HIPAA's security mandates when storing health information.²⁰⁸ Thus, expansion of HIPAA could provide patients with valuable protections against disclosure and misuse of their health information to other stakeholders with an interest in predicting an individual's future health.

203. *Id.*

204. Hoffman, *Big Data and the ADA*, *supra* note 76, at 779.

205. Andrew Marder, *How Does Life Insurance Work?*, NERDWALLET (July 13, 2020), <https://www.nerdwallet.com/blog/insurance/how-does-life-insurance-work> [<https://perma.cc/W3CR-RCH6>].

206. Hoffman, *Genetic Testing & PHA Regulation*, *supra* note 48, at 156–58.

207. *Id.* at 157.

208. *Id.* at 158.

CONCLUSION

Predictive analytics in healthcare presents a promising solution for improving patient outcomes and reducing healthcare costs. But, like genetic testing, predictive analytics has generated both excitement and concerns. The objective of this Note is not to hinder predictive analytics research; rather, this Note concludes that legislative protections may be needed to assure patients that their healthcare information (and other data) will not be used to discriminate against them based on their future health status. Without such protections, individuals may be wary of predictive analytics models or be reluctant to consent, if given the opportunity to withhold consent, to their data being used in such models. The question is not *if*, but *when* the time will come to address concerns regarding disclosure of predicted health outcomes to patients and discrimination based on future health status. Current data privacy and antidiscrimination laws may not adequately protect individuals against discrimination by parties interested in their future health status. Providers must also anticipate how they will communicate, and which results from predictive health analytics models they should disclose to patients. As predictive analytics models advance from pilots and research to clinical and commercial use, the legal and ethical responses that emerge will play a crucial role in shaping the effects that predictive health analytics has on the U.S. healthcare system.