

WHAT GENETIC TESTING TEACHES ABOUT PREDICTIVE HEALTH ANALYTICS REGULATION*

SHARONA HOFFMAN**

The ever-growing phenomenon of predictive health analytics is generating significant excitement, hope for improved health outcomes, and potential for new revenues. Researchers are developing algorithms to predict suicide, heart disease, stroke, diabetes, cognitive decline, opioid abuse, cancer recurrence, and other ailments. The researchers include not only medical experts, but also commercial enterprises, such as Facebook and LexisNexis, who may profit from the work considerably. This Article focuses on long-term disease predictions (i.e., predictions regarding future illnesses), which have received surprisingly little attention in the legal and ethical literature. It compares the robust academic and policy debates and legal interventions that followed the emergence of genetic testing to the relatively anemic reaction to predictions produced by artificial intelligence and other predictive methods. This Article argues that, like genetic testing, predictive health analytics raises significant concerns about psychological harm, privacy breaches, discrimination, and the meaning and accuracy of predictions. Consequently, as alluring as the new predictive technologies are, they require careful consideration and thoughtful safeguards. These include changes to the HIPAA Privacy and Security Rules and the Americans with Disabilities Act, careful oversight mechanisms, and self-regulation by healthcare providers. Ignoring the hazards of long-term predictive health analytics and failing to provide data subjects with appropriate rights and protections would be a grave mistake.

INTRODUCTION 124
I. LONG-TERM PREDICTIVE HEALTH ANALYTICS..... 127

* © 2019 Sharona Hoffman.

** Edgar A. Hahn Professor of Law and Professor of Bioethics, Co-Director of Law-Medicine Center, Case Western Reserve University School of Law; B.A., Wellesley College; J.D., Harvard Law School; LL.M. in Health Law, University of Houston; S.J.D. in Health Law, Case Western Reserve University. Author of *Electronic Health Records and Medical Big Data: Law and Policy* (Cambridge Univ. Press 2016). A huge thank you to Computer Science Professor Andy Podgurski for everything he taught me about predictive health analytics and for his astute comments on prior drafts. Thank you also to Professors Anant Madabhushi and Max Mehlman for their very helpful suggestions and to Melissa Vogley for her skillful research assistance.

124		[Vol. 98
	<i>NORTH CAROLINA LAW REVIEW</i>	
	A. <i>Predictive Health Analytics Defined</i>	127
	B. <i>Long-Term Predictive Health Analytics Examples</i>	130
	C. <i>Nontraditional Data Sources for Predictive Health Analytics</i> ...	132
	D. <i>Predictive Health Analytics as Big Business</i>	134
II.	THE PRECEDENT OF GENETIC TESTING.....	135
	A. <i>Genetic Testing</i>	136
	B. <i>Genetic Testing Concerns</i>	137
	1. <i>Clinical Validity and Accuracy</i>	137
	2. <i>Privacy and Discrimination</i>	138
	3. <i>Psychological Harms</i>	140
	C. <i>Legal and Policy Interventions</i>	142
	1. <i>State and Federal Antidiscrimination Legislation</i>	142
	2. <i>The HIPAA Privacy Rule</i>	144
	3. <i>Self-Regulation</i>	146
III.	LONG-TERM PREDICTIVE HEALTH ANALYTICS CONCERNS	147
	A. <i>Psychological Harms</i>	148
	B. <i>Privacy and Discrimination</i>	149
	C. <i>Erroneous Predictions</i>	151
	D. <i>Regulatory Uncertainty</i>	154
IV.	RECOMMENDATIONS.....	156
	A. <i>Legal Interventions</i>	156
	1. <i>The HIPAA Privacy and Security Rules</i>	156
	2. <i>The Americans with Disabilities Act</i>	158
	B. <i>Other Oversight Mechanisms</i>	159
	1. <i>Guidelines and Validation</i>	159
	2. <i>Self-Regulation</i>	161
	CONCLUSION.....	162

INTRODUCTION

The ever-growing phenomenon of predictive health analytics¹ is generating significant excitement, hope for improved health outcomes, and potential for new revenues.² Researchers are developing algorithms to predict suicide, heart disease, stroke, diabetes, cognitive decline, opioid abuse, cancer

1. See *infra* notes 19–24 and accompanying text for a definition and discussion of predictive health analytics.

2. See Jennifer Bresnick, *10 High-Value Use Cases for Predictive Analytics in Healthcare*, HEALTH IT ANALYTICS (Sept. 4, 2018), <https://healthitanalytics.com/news/10-high-value-use-cases-for-predictive-analytics-in-healthcare> [<https://perma.cc/S4SQ-HDGM>].

recurrence, and other ailments.³ In 2017, the Society of Actuaries found that ninety-three percent of healthcare and health insurance executives surveyed believed that predictive analytics is important to their future success.⁴ Indeed, some experts forecast that predictive health analytics will be a commonplace medical tool in the near future.⁵

Healthcare providers can also use predictive health analytics for treatment purposes in the short term.⁶ For example, predictive health analytics can help physicians identify patients who are at risk of hospital readmission because of complications.⁷ This Article, however, focuses on health analytics that predicts health problems in the more distant future, which this Article calls “long-term predictive health analytics.” For instance, scientists are developing techniques to forecast conditions such as heart disease or cognitive decline that could be years or decades away.⁸

In some instances, such forecasts can be medically beneficial because they are clinically actionable—clinicians can commence early screening of known affected individuals and implement preventive interventions.⁹ In the case of heart disease, for example, these actions might include drugs, exercise, and improved diet.¹⁰

At the same time, predictive health analytics can be potentially harmful.¹¹ Individuals who are identified as having a high risk of developing future health problems, such as cognitive decline or opioid addiction, may suffer psychological distress, privacy violations (if the information is circulated to unauthorized third parties), discrimination, and other harms.¹² One scholar has worried that people labelled as being at high risk of suicide will be treated

3. See *infra* Section I.B.

4. SOC'Y OF ACTUARIES, PREDICTIVE ANALYTICS IN HEALTHCARE TREND FORECAST 2, 4 (2017), <https://www.soa.org/Files/programs/predictive-analytics/2017-health-care-trend.pdf> [<https://perma.cc/A73Q-ZVA6>]. The survey included 223 participants. *Id.* at 2.

5. See Eric J. Topol, *High-Performance Medicine: The Convergence of Human and Artificial Intelligence*, 25 NATURE MED. 44, 44 (2019).

6. I. Glenn Cohen et al., *The Legal and Ethical Concerns That Arise from Using Complex Predictive Analytics in Health Care*, 33 HEALTH AFF. 1139, 1140 (2014) (explaining that “it has become possible to apply predictive analytics to health care”).

7. *Id.*

8. See *infra* Section I.B.

9. Bresnick, *supra* note 2.

10. *Strategies to Prevent Heart Disease*, MAYO CLINIC (Jan. 9, 2019), <https://www.mayoclinic.org/diseases-conditions/heart-disease/in-depth/heart-disease-prevention/art-20046502> [<https://perma.cc/FKV2-FL7E>].

11. See *infra* Part III.

12. See *infra* Part III.

differently by their physicians.¹³ Specifically, physicians might discontinue beneficial medications for fear that they will exacerbate the suicide risk, unnecessarily send police to patients' homes, forcibly hospitalize patients, or relate to them in a demeaning, dehumanizing way.¹⁴

Moreover, predictive health analytics outcomes can be erroneous for a variety of reasons.¹⁵ Thus, individuals may endure serious adverse consequences based on mistaken predictions when, in truth, there is no evidence they are at risk of developing the alleged health problems.

This Article argues that we are doing alarmingly little to identify and address the ethical and legal implications of long-term predictive health analytics. This is in stark contrast to policymakers' thoughtful approach to the emergence of genetic testing several decades ago.¹⁶ This Article highlights the discrepancy between society's relatively cautious approach to genetic testing and its more cavalier approach to predictive analytics and argues that, as they did in the case of genetic testing, scientists must carefully consider the benefits and risks of predictive health analytics and implement safeguards to address its potential hazards.

Accordingly, this Article suggests that data subjects should enjoy rights that give them some degree of control over their data including predicted health outcomes. They should have expanded rights to consent to disclosure of their health information, to discover who has seen their health data, and to sue for both privacy breaches that harm them and for discrimination based on disease predictions.¹⁷ Additionally, the scientific community should develop oversight mechanisms to safeguard the quality of predictive models.¹⁸

The remainder of this Article proceeds as follows: Part I describes long-term predictive health analytics and illustrates the work that scientists are conducting in this area; Part II analyzes the precedent of genetic testing, focusing on the concerns that it raised and the measures that policymakers implemented to address those concerns; Part III examines the risks of long-term predictive health analytics; Part IV develops preliminary recommendations for responsive legal and policy changes; and finally, Part V concludes.

13. See Mason Marks, *Artificial Intelligence Based Suicide Prediction*, YALE J. HEALTH POL'Y L. & ETHICS (forthcoming 2019) (manuscript at 22) (on file with North Carolina Law Review) [hereinafter Marks, *Artificial Intelligence*] ("People placed in this category [of suicide-prone individuals] may be treated differently by physicians in ways that endanger their health and safety.").

14. *Id.* (manuscript at 22–24).

15. See *infra* Section III.C.

16. See *infra* Part II.

17. See *infra* Section IV.A.

18. See *infra* Section IV.B.

I. LONG-TERM PREDICTIVE HEALTH ANALYTICS

A. *Predictive Health Analytics Defined*

Predictive analytics is defined as “the analysis of large data sets to discover patterns and [the] use [of] those patterns to forecast or predict the likelihood of future events.”¹⁹ Experts conduct this analysis using computer algorithms.²⁰ An algorithm is a precise step-by-step process that leaves nothing to guesswork or intuition.²¹ Learning algorithms train predictive models using training sets comprised of sample input and output values.²² Some analysts use the term “predictive modeling,” which can be defined as “the process of developing a mathematical tool or model that generates an accurate prediction.”²³ Researchers often use the terms “learning algorithm” and “predictive model” interchangeably, although the term “predictive model” suggests a representation of knowledge that is created by an algorithm.²⁴ Predictive analytics is based on techniques from three closely related areas of research: statistical inference, data mining, and machine learning.²⁵

Statistical inference involves analyzing a sample dataset, inferring properties of a larger population based on this sample, and characterizing uncertainties about those properties.²⁶ Data mining is “the practice of searching through large amounts of computerized data to find useful patterns or trends.”²⁷

19. David Crockett, Ryan Johnson & Brian Eliason, *What is Data Mining in Healthcare?*, HEALTH CATALYST 1, 1 (2017), <https://www.healthcatalyst.com/wp-content/uploads/2014/06/What-is-data-mining-in-healthcare.pdf> [<https://perma.cc/8PUQ-F6CY>].

20. See Cohen et al., *supra* note 6, at 1139; W. Nicholson Price II, *Regulating Black-Box Medicine*, 116 MICH. L. REV. 421, 425–26 (2017) (discussing the nature of medical algorithms).

21. Deven R. Desai & Joshua A. Kroll, *Trust but Verify: A Guide to Algorithms and the Law*, 31 HARV. J.L. & TECH. 1, 23 (2017).

22. See SHAI SHALEV-SHWARTZ & SHAI BEN-DAVID, UNDERSTANDING MACHINE LEARNING: FROM THEORY TO ALGORITHMS 13–14 (2014) (discussing “the statistical learning framework”).

23. MAX KUHN & KJELL JOHNSON, APPLIED PREDICTIVE MODELING 2 (2013).

24. See *supra* notes 22–23 and accompanying text.

25. See generally BRUCE RATNER, STATISTICAL AND MACHINE-LEARNING DATA MINING: TECHNIQUES FOR BETTER PREDICTIVE MODELING AND ANALYSIS OF BIG DATA 3–11 (3d ed. 2017) (ebook) (discussing statistical inference, data mining, and machine learning).

26. WILLIAM L. HAYS, STATISTICS 1 (4th ed. 1988) (describing statistical inference as a process of analysis that enables one to “make general statements about the large body of potential observations, of which the data collected represents but a sample”); *Statistical Inference*, OXFORD DICTIONARY, https://en.oxforddictionaries.com/definition/us/statistical_inference [<https://perma.cc/2VLV-H48N>] (defining statistical inference as “[t]he theory, methods, and practice of forming judgments about the parameters of a population and the reliability of statistical relationships, typically on the basis of random sampling”).

27. *Data Mining*, MERRIAM-WEBSTER, <https://www.merriam-webster.com/dictionary/data%20mining> [<https://perma.cc/9Y3R-J2W6>].

In more technical terms, it is the process of using algorithms to examine “big data” from sources such as databases or the internet in order to unearth hidden knowledge or patterns.²⁸ “Big data” is characterized by its “three Vs”: high volume, variety, and velocity, the last referring to the speed with which data are generated.²⁹ In healthcare, big data can come from a myriad of sources, including patients, healthcare providers, insurers, manufacturers, the government, and even mobile devices such as smartphones and wearables.³⁰ For example, Geisinger Health System, which operates in Pennsylvania and New Jersey, has created a large unified data architecture.³¹ It draws upon a variety of sources including electronic health records, patient satisfaction surveys, and wellness apps.³²

Readers are most likely familiar with the general term “artificial intelligence,” which refers to computers’ ability to mimic human behavior and learn.³³ One type of artificial intelligence is machine learning, which refers to methods that enable computers to “automatically detect patterns in data, and then use the uncovered patterns to predict future data, or to perform other kinds of decision-making under uncertainty.”³⁴ Scientists train computers to do analytical work by feeding them information, such as patients’ medical

28. JIAWEI HAN, MICHELINE KAMBER & JIAN PEI, *DATA MINING: CONCEPTS AND TECHNIQUES* 8 (3d ed. 2012) (ebook).

29. SHARONA HOFFMAN, *ELECTRONIC HEALTH RECORDS AND MEDICAL BIG DATA: LAW AND POLICY* 111 (2016); W. Nicholson Price II & I. Glenn Cohen, *Privacy in the Age of Medical Big Data*, 25 *NATURE MED.* 37, 37 (2019).

30. Nathan Cortez, *Substantiating Big Data in Health Care*, 14 *I/S* 61, 63–64 (2017) (discussing the breadth of big data sources).

31. See generally Alistair R. Erskine et al., *How Geisinger Health System Uses Big Data to Save Lives*, *HARV. BUS. REV.* (Dec. 15, 2016), <https://hbr.org/2016/12/how-geisinger-health-system-uses-big-data-to-save-lives> [<https://perma.cc/W9QN-4T4A>] (describing the Geisinger data system, its creation, and its uses).

32. *Id.* (explaining that the system integrates data, with patient permission, reports from different departmental systems and hospitals, information from participation surveys, and data from wellness apps to get a more complete understanding of the patient’s health).

33. See IAN GOODFELLOW, YOSHUA BEBGIO & AARON COURVILLE, *DEEP LEARNING* 1–8 (2016) (discussing the basic notion of artificial intelligence and the different learning process that can be applied to create AI).

34. KEVIN P. MURPHY, *MACHINE LEARNING: A PROBABILISTIC PERSPECTIVE* 1 (2012); see also David Lehr & Paul Ohm, *Playing with the Data: What Legal Scholars Should Learn About Machine Learning*, 51 *U.C. DAVIS L. REV.* 653, 671 (2017) (“Fundamentally, machine learning refers to an automated process of discovering correlations (sometimes alternatively referred to as relationships or patterns) between variables in a dataset, often to make predictions or estimates of some outcome.”); Alvin Rajkomar, Jeffrey Dean & Isaac Kohane, *Machine Learning in Medicine*, 380 *NEW ENG. J. MED.* 1347, 1348 (2019) (explaining that “in machine learning, a model learns from examples rather than being programmed with rules”).

records.³⁵ For example, scientists might show computers a large number of tumor images with indications as to which ones are cancerous and which ones are not.³⁶ The computers then learn to differentiate between benign and malignant tumors based on patterns in the tumor x-rays or scans so that they can identify cancerous tumors when shown new images.³⁷

A well-known type of machine learning is deep learning, which allows computers “to learn from experience and understand the world in terms of a hierarchy of concepts, with each concept defined through its relation to simpler concepts.”³⁸ Thus, computers gather knowledge from experience and learn more complex concepts by building on simpler concepts.³⁹

Predictive models are valuable for physicians, researchers, and policymakers.⁴⁰ They can help public health officials identify those who are at highest risk of developing a disease so that health officials can implement preventive interventions.⁴¹ In the clinical setting, predictive models may discern which patients are likely to have poor or successful treatment outcomes so physicians can tailor their medical decisions accordingly.⁴² Predictive analytics may also help identify high-risk individuals whom doctors should aggressively screen for particular diseases.⁴³

Thus, predictive health analytics can generate projections of health problems that may plague individuals in the future.⁴⁴ Such predictions can be beneficial to patients if physicians intervene to prevent or detect the condition at a very early stage. However, such predictions can also render the patient

35. See Niha Beig et al., *Perinodular and Intranodular Radiomic Features on Lung CT Images Distinguish Adenocarcinomas from Granulomas*, 290 *RADIOLOGY* 783, 784 (2019) (relating that a “machine classifier was trained on a cohort of 145 patients”).

36. *See id.*

37. *See id.* at 792.

38. GOODFELLOW ET AL., *supra* note 33, at 1.

39. *Id.* For example, the technique is used for facial recognition. See Divyansh Dwivedi, *Face Detection for Beginners*, TOWARDS DATA SCI. (Apr. 27, 2018), <https://towardsdatascience.com/face-detection-for-beginners-e58e8f21aad9> [<https://perma.cc/QM8P-TCU3>]. A deep learner might first recognize simple features such as lines, then slightly more complex features such as regions, and finally entire faces. *See id.*; see generally Wenzhi Zhao & Shihong Du, *Spectral–Spatial Feature Extraction for Hyperspectral Image Classification: A Dimension Reduction and Deep Learning Approach*, 54 *IEEE TRANSACTIONS ON GEOSCIENCE & REMOTE SENSING* 4544, 4544 (2016) (describing deep learning approaches to using spectral–spatial feature classifications).

40. EWOUT W. STEYERBERG, *CLINICAL PREDICTION MODELS* 1–2, 11 (2009).

41. *Id.* at 11–12.

42. *Id.* at 11.

43. *Id.* at 11–12.

44. *See infra* Section I.B.

vulnerable to adverse psychological consequences, discrimination, and other harms.⁴⁵

B. *Long-Term Predictive Health Analytics Examples*

Scientists are working hard to identify physical and behavioral clues that might indicate an individual's future health status. Many studies focus on the question of whether there are traits, habits, or other indicators that signal that an individual is vulnerable to particular diseases in the future.⁴⁶

Currently, medical researchers are investigating biomarkers that help them discern disease risks. A "biomarker" is a "biological molecule found in blood, other body fluids, or tissues that is a sign of a normal or abnormal process, or of a condition or disease."⁴⁷ For example, researchers reported in 2014 that people with lower levels of ten identified phospholipids in their blood were at higher risk of having existing cognitive impairments or declining cognitively within a few years.⁴⁸

Other physiological phenomena can also serve as predictors of future health risks. A 2018 study focused on retinopathy,⁴⁹ an eye condition that is caused by damage to the small retinal blood vessels in the eye,⁵⁰ found that retinopathy was associated with higher rates of cognitive decline over the next twenty years.⁵¹

Human eyes can also reveal information about cardiovascular risks. In 2018, researchers from Google and its health-tech subsidiary, Verily, used carefully validated deep learning models trained on medical data from nearly 300,000 patients to predict the values of known heart disease risk factors from

45. See *infra* Part III.

46. See Ran Balicer, *The Doctor Will See Your Future Now*, FORBES (Apr. 16, 2018), <https://www.forbes.com/sites/startupnationcentral/2018/04/16/for-predictive-medicine-its-back-to-the-future/#5d1714da3525> [<https://perma.cc/XH29-FEU7>].

47. NCI *Dictionary of Cancer Terms*, NAT'L CANCER INST., www.cancer.gov/publications/dictionaries/cancer-terms?cdrid=45618 [<https://perma.cc/RQP3-CLLR>].

48. Alison Abbott, *Biomarkers Could Predict Alzheimer's Before It Starts*, NATURE (Mar. 9, 2014), <https://www.nature.com/news/biomarkers-could-predict-alzheimer-s-before-it-starts-1.14834> [<https://perma.cc/7FCS-E22L>].

49. See Jennifer A. Deal et al., *Retinal Signs and 20-Year Cognitive Decline in the Atherosclerosis Risk in Communities Study*, 90 NEUROLOGY e1158, e1158 (2018).

50. Janet M. Torpy, *Retinopathy*, 298 JAMA 944, 944 (2007); *Diabetic Retinopathy*, MAYO CLINIC, <https://www.mayoclinic.org/diseases-conditions/diabetic-retinopathy/symptoms-causes/syc-20371611> [<https://perma.cc/3LGV-NRRE>].

51. Deal et al., *supra* note 49, at e1158, e1165. The study involved 12,317 men and women who were fifty to seventy-three years of age when they were first examined. *Id.*

retinal fundus images.⁵² The risk factors included age, gender, smoking status, systolic blood pressure, and past major cardiac events.⁵³ The deep learner then used images to predict heart disease risks directly.⁵⁴ In addition, the researchers identified the anatomical regions the deep learner might have been using to make its predictions, which they believed included the optic nerve and blood vessels.⁵⁵ The authors acknowledged that further research with larger datasets is needed to verify their results.⁵⁶

IBM researchers “identified an automated machine-learning speech classifier” that could predict psychosis based on the speech patterns of high-risk patients.⁵⁷ The technique relied on indicators such as less semantic coherence and diminished use of possessive pronouns and reportedly achieved an eighty-three percent accuracy rate.⁵⁸

Learning algorithms have been able to predict Alzheimer’s disease up to six years before it manifests.⁵⁹ Researchers are also working to determine whether machine learning can predict premature death associated with chronic disease.⁶⁰

Oncologists have developed algorithms to predict patients’ prognosis after cancer treatment.⁶¹ For example, they have developed machine-learning tools

52. Ryan Poplin et al., *Prediction of Cardiovascular Risk Factors from Retinal Fundus Photographs via Deep Learning*, 2 NATURE BIOMEDICAL ENGINEERING 158, 158 (2018); James Vincent, *Google’s New AI Algorithm Predicts Heart Disease by Looking at Your Eyes*, VERGE (Feb. 19, 2018), <https://www.theverge.com/2018/2/19/17027902/google-verily-ai-algorithm-eye-scan-heart-disease-cardiovascular-risk> [https://perma.cc/P4KB-RX34] (“As with all deep learning analysis, neural networks were then used to mine this information for patterns, learning to associate telltale signs in the eye scans with the metrics needed to predict cardiovascular risk (e.g., age and blood pressure).”). The retinal fundus is the interior lining of the eyeball, including the retina, optic disc, and the macula. William C. Shiel Jr., *Medical Definition of Retinal Fundus*, MEDICINENET (Dec. 27, 2018), <https://www.medicinenet.com/script/main/art.asp?articlekey=10601> [https://perma.cc/3RWG-BYCW].

53. Poplin et al., *supra* note 52, at 158, 161–62.

54. *Id.*

55. *Id.*

56. *Id.* at 161–62.

57. Cheryl M. Corcoran et al., *Prediction of Psychosis Across Protocols and Risk Cohorts Using Automated Language Analysis*, 17 WORLD PSYCHIATRY 67, 67 (2018).

58. *Id.* at 67, 70.

59. Ana Sandoiu, *Using Artificial Intelligence to Predict Mortality*, MED. NEWS TODAY (Mar. 29, 2019), <https://www.medicalnewstoday.com/articles/324828.php> [https://perma.cc/UN2J-S5W9].

60. *Id.*

61. See Masahiro Takada et al., *Prediction of Postoperative Disease-Free Survival and Brain Metastasis for HER2-Positive Breast Cancer Patients Treated with Neoadjuvant Chemotherapy Plus Trastuzumab Using a Machine Learning Algorithm*, 172 BREAST CANCER RES. & TREATMENT 611, 611 (2018).

to predict the likelihood of recurrence and brain metastasis in certain breast cancer and lung cancer patients.⁶²

Electronic documentation has been particularly helpful for purposes of health predictions. In 2014, IBM announced that it had analyzed electronic health records from Virginia's Carilion Clinic and was able to identify 8500 patients who were at risk of heart failure.⁶³ Scientists have also been able to analyze electronic health records and medical claims data to predict which individuals will develop depression or diabetes-related problems up to a year in advance.⁶⁴ The U.S. Department of Veteran Affairs recently launched a program called "VA REACH VET" that uses a predictive model to analyze veterans' electronic health records and identify individuals at high risk of suicide.⁶⁵

C. *Nontraditional Data Sources for Predictive Health Analytics*

Analysts are turning to nontraditional data sources as well. For example, several years ago, Carolinas Healthcare (now Atrium Health) purchased consumer information from data brokers⁶⁶ in an effort to use algorithms to identify high-risk patients.⁶⁷ Information garnered from credit card purchasing

62. *Id.* at 611, 616. See generally Jing Zhong et al., *Constructing a Risk Prediction Model for Lung Cancer Recurrence by Using Gene Function Clustering and Machine Learning*, 22 COMBINATORIAL CHEMISTRY & HIGH THROUGHPUT SCREENING 226 (2019) (discussing, in the case of lung cancer, the use of preventive screening tools to identify risk of metastasis in patients and to intervene earlier).

63. Mohana Ravindranath, *IBM Used Predictive Analytics to Find Patients at Risk of Heart Failure*, WASH. POST (Feb. 20, 2014), www.washingtonpost.com/business/on-it/ibm-used-predictive-analytics-to-find-patients-at-risk-of-heart-failure/2014/02/20/9b0ddb3c-9a47-11e3-b88d-f36c07223d88_story.html [<https://perma.cc/JM6G-JJAQ> (dark archive)]; Press Release, IBM, IBM Predictive Analytics to Detect Patients at Risk for Heart Failure (Feb. 19, 2014), <http://www-03.ibm.com/press/us/en/pressrelease/43231.wss> [<https://perma.cc/2CHV-UDTY>].

64. Arthur Allen, *Big Brother Is Watching Your Waist*, POLITICO (July 21, 2014), www.politico.com/story/2014/07/data-mining-health-care-109153 [<https://perma.cc/7RBA-YH6D>].

65. Press Release, U.S. Dep't Veterans Affairs, VA REACH VET Initiative Helps Save Veterans Lives: Program Signals When More Help Is Needed for At-Risk Veterans (Apr. 3, 2017), <https://www.va.gov/opa/pressrel/includes/viewpdf.cfm?id=2878> [<https://perma.cc/PV2U-E8FX>].

66. Data brokers are "companies that collect information, including personal information about consumers, from a wide variety of sources for the purpose of reselling such information to their customers for various purposes . . ." FED. TRADE COMM'N, PROTECTING CONSUMER PRIVACY IN AN ERA OF RAPID CHANGE 68 (2012), <https://www.ftc.gov/sites/default/files/documents/reports/federal-trade-commission-report-protecting-consumer-privacy-era-rapid-change-recommendations/120326privacyreport.pdf> [<https://perma.cc/T8SJ-P2GA>]; see also Janine S. Hiller, *Healthy Predictions? Questions for Data Analytics in Health Care*, 53 AM. BUS. L.J. 251, 271–72 (2016) (discussing data brokers and health data streams).

67. Melanie Hicken, *Big Data: Look Who's Buying Your Personal Information*, CNN MONEY (Sept. 10, 2014), <https://money.cnn.com/gallery/pf/2014/09/07/big-data-personal-information/3.html> [<https://perma.cc/SNF5-N7HC>]; Shannon Pettypiece & Jordan Robertson, *Hospitals, Including Carolinas HealthCare, Using Consumer Purchase Data for Information on Patient Health*, CHARLOTTE

records or grocery loyalty cards can indicate whether individuals are buying healthy food, smoking, refilling their prescriptions, and buying gym memberships.⁶⁸ These data points in turn can predict the likelihood that someone will have a severe asthma or heart attack.⁶⁹

The University of Pittsburgh Medical Center has used patient demographic and household information to predict health risks.⁷⁰ It concluded that people who do not reside with children and earn less than \$50,000 annually are more likely to go to an emergency room rather than visit a doctor's office, even though the latter is a much less costly option that is appropriate for many conditions.⁷¹ Likewise, it concluded that individuals without a car may not be receiving adequate medical care.⁷² Healthcare systems assert that they use such information in order to implement preventive and corrective interventions for patients.⁷³ However, skeptics have questioned their true motivations, suspecting that cost savings are at the heart of the matter and worrying that data mining practices compromise patient privacy and damage the physician-patient relationship.⁷⁴

Social media has become an increasingly common source of data used for predictive health analytics as well. Researchers recently reported that they used an algorithm to analyze Facebook data from close to 1200 consenting users and identified linguistic signals that could predict future depression.⁷⁵

Facebook itself has joined the fray of predictive health analytics. Its software now monitors users' posts to identify those with suicidal intent, and an algorithm assigns a risk score ranging from zero to one.⁷⁶ The algorithm interprets phrases such as "Are you okay?" paired with "Goodbye" and "Please

OBSERVER (June 27, 2014), <https://www.charlotteobserver.com/living/health-family/article9135980.html> [<https://perma.cc/V69S-TK4U> (staff uploaded archive)].

68. Pettypiece & Robertson, *supra* note 67.

69. *Id.*

70. *Id.*

71. *Id.*

72. *See id.*

73. *See id.*

74. *See id.* For a discussion of concerns raised by predictive analytics, see *infra* Part III.

75. Johannes C. Eichstaedt et al., *Facebook Language Predicts Depression in Medical Records*, 115 PROC. NAT'L ACAD. SCI. 11203, 11203, 11207 (2018).

76. 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), <https://www.businessinsider.com/facebook-is-using-ai-to-try-to-predict-if-youre-suicidal-2018-12> [<https://perma.cc/5KBE-MG8T>]; Martin Kaste, *Facebook Increasingly Reliant on A.I. To Predict Suicide Risk*, NPR (Nov. 17, 2018), <https://www.npr.org/2018/11/17/668408122/facebook-increasingly-reliant-on-a-i-to-predict-suicide-risk> [<https://perma.cc/R932-9JPL>].

don't do this" as clues that someone is in distress.⁷⁷ In cases it assesses as severe, Facebook contacts the police, as it did at least 3500 times in 2018.⁷⁸ Unfortunately, police officers who are poorly trained or inexperienced may mishandle such "wellness checks," exacerbating the situation and, in extreme cases, using deadly force against individuals with mental illness.⁷⁹ Consequently, predictions regarding suicide can, ironically, be dangerous and harmful for data subjects.

D. *Predictive Health Analytics as Big Business*

Predictive health analytics has already generated business opportunities for enterprising organizations. Companies are reportedly selling "risk scores" to health care providers and insurers to identify patients who are at risk of becoming addicted to or overdosing on opioids.⁸⁰ Business giants such as LexisNexis collect data from insurance claims, electronic health records, housing information, and records relating to patients' social and family connections in order to produce risk scores.⁸¹ They do all of this without asking patients for permission and are not required to seek consent by law.⁸²

Data brokers sell other types of information to health care providers as well. For example, LexisNexis and Acxiom sell assessments of patients based on "criminal records, online purchasing histories, retail loyalty programs and voter

77. Mason Marks, *Suicide Prediction Technology Is Revolutionary. It Badly Needs Oversight*, WASH. POST (Dec. 20, 2018), https://www.washingtonpost.com/outlook/suicide-prediction-technology-is-revolutionary-it-badly-needs-oversight/2018/12/20/214d2532-fd6b-11e8-ad40-cdfd0e0dd65a_story.html?utm_term=.951089ae3f76 [<https://perma.cc/SAK5-QCQ2> (dark archive)].

78. Kaste, *supra* note 76 (explaining that Facebook software first scans individuals' accounts for indications of "imminent self-harm," then flags individuals for Facebook employees, who decide whether or not to alert the police).

79. See Marks, *Artificial Intelligence*, *supra* note 13 (manuscript at 24).

80. Mohana Ravindranath, *How Your Health Information Is Sold and Turned into 'Risk Scores'*, POLITICO (Feb. 3, 2019), <https://www.politico.com/story/2019/02/03/health-risk-scores-opioid-abuse-1139978> [<https://perma.cc/MYM8-2V48>] [hereinafter Ravindranath, *Risk Scores*]; see also Timothy R. Hylan et al., *Automated Prediction of Risk for Problem Opioid Use in a Primary Care Setting*, 16 J. PAIN 380, 385 (2015) (discussing use of electronic health records to develop "simple risk stratification algorithms to initially alert clinicians to . . . patients at higher risk for problem opioid use").

81. Ravindranath, *Risk Scores*, *supra* note 80.

82. *Id.* For a discussion of the limitations of the HIPAA Privacy Rule, see *infra* notes 167–68 and accompanying text.

registration data.”⁸³ This information is used to identify individuals who are at risk of requiring costly care or readmission to a hospital.⁸⁴

Moreover, data brokers routinely supply predictive health information to parties outside the healthcare industry. They garner data from a myriad of sources, such as publicly available records, surveys, shopper loyalty programs, social media, magazine subscription lists, fitness devices, people’s internet searches, and more.⁸⁵ They then organize and sell the data, often with personally identifying information, to interested third parties, including marketers.⁸⁶ These entities in turn can then use the medical information for the purpose of predictive analytics: to predict individuals’ future behaviors and health needs.⁸⁷

II. THE PRECEDENT OF GENETIC TESTING

Predictive health analytics is novel and exciting, but it is not the first mechanism used to predict future health problems. A much more familiar and well-established technique is genetic testing, also known as DNA testing.⁸⁸ When genetic testing emerged as a prevalent diagnostic and predictive tool, it raised significant ethical, legal, and policy concerns.⁸⁹ There is much to be learned from the conversations and interventions that followed. This part provides background information regarding genetic testing, analyzes the

83. Mohana Ravindranath, *Does Your Doctor Need to Know What You Buy on Amazon?*, POLITICO (Oct. 30, 2018), <https://www.politico.com/story/2018/10/30/the-doctor-will-see-through-you-now-893437> [<https://perma.cc/5V7N-SDKN>].

84. *Id.* For a discussion of the privacy and discrimination concerns that such practices raise, see *infra* Section III.B.

85. Adam Tanner, *Strengthening the Protection of Patient Medical Data*, CENTURY FOUND. (Jan. 10, 2017), <https://tcf.org/content/report/strengthening-protection-patient-medical-data/?agreed=1> [<https://perma.cc/T2J9-KP8K> (staff uploaded archive)]; Sam Thielman, *Your Private Medical Data Is for Sale – and It’s Driving a Business Worth Billions*, GUARDIAN (Jan. 10, 2017), <https://www.theguardian.com/technology/2017/jan/10/medical-data-multibillion-dollar-business-report-warns> [<https://perma.cc/2EMX-DZYE>].

86. See Tanner, *supra* note 85.

87. Russ Cobb, 2018: *Taking Your Healthcare System Marketing Strategies into the Consumer Age*, BECKER’S HOSP. REV. (Dec. 19, 2017), <https://www.beckershospitalreview.com/patient-engagement/2018-taking-your-healthcare-system-marketing-strategies-into-the-consumer-age.html> [<https://perma.cc/QDZ4-K2UZ>]; see Thielman, *supra* note 85. For a discussion of the privacy and discrimination concerns that such practices raise, see *infra* Section III.B.

88. See Wylie Burke, *Genetic Testing*, 347 NEW ENG. J. MED. 1867, 1867 (2002); *Genetic Testing FAQ*, NAT’L HUM. GENOME RES. INST., <https://www.genome.gov/19516567/faq-about-genetic-testing/> [<https://perma.cc/AH3F-9E2C>] (“Genetic testing uses laboratory methods to look at your genes, which are the DNA instructions you inherit from your mother and your father. . . . Genetic tests may be used to identify increased risks of health problems, to choose treatments, or to assess responses to treatments.”); *Genetic Testing: How It Is Used for Healthcare*, U.S. DEP’T HEALTH & HUM. SERVS., <https://report.nih.gov/nihfactsheets/ViewFactSheet.aspx?csid=43> [<https://perma.cc/DE2G-WQ5U>].

89. See *infra* Section II.B.

concerns that it raised, and discusses the legal and policy interventions that were implemented to address some of those concerns. Many parallels can be drawn between genetic testing and predictive health analytics. However, thus far, there has been little outcry for safeguards relating to long-term health predictions that are not based on genetics.

A. Genetic Testing

In the late 1960s, scientists developed the ability to test fetuses for Down syndrome with a sample of amniotic fluid.⁹⁰ Fetal genetic testing became common beginning in the 1970s, and today, it is used to screen for Tay-Sachs disease, sickle cell disease, cystic fibrosis, and many other illnesses.⁹¹

Genetic testing can also analyze disease risks after birth and provide information regarding the likelihood that individuals will develop specific maladies in the future.⁹² In 1990, Mary King-Claire identified a genetic mutation,⁹³ BRCA1, that is linked to breast and ovarian cancer, as is BRCA2, which was discovered shortly thereafter.⁹⁴ Since then, scientists have discovered a myriad of genetic abnormalities that can increase disease vulnerabilities and have developed predictive genetic tests for some of them.⁹⁵ For example, predictive testing can be done for early-onset familial Alzheimer's disease, a

90. Glenn E. Palomaki, *Screening for Down's Syndrome*, 333 NEW ENG. J. MED. 532, 532 (1995) (book review). Amniotic fluid is the fluid that surrounds the fetus in the womb. Lori Smith, *What's to Know About Amniotic Fluid?*, MED. NEWS TODAY (June 27, 2018), <https://www.medicalnewstoday.com/articles/307082.php> [<https://perma.cc/7P67-K64F>].

91. Nancy Press, *Genetic Testing and Screening*, in FROM BIRTH TO DEATH AND BENCH TO CLINIC: THE HASTINGS CENTER BIOETHICS BRIEFING BOOK FOR JOURNALISTS, POLICYMAKERS, AND CAMPAIGNS 73, 73 (Mary Crowley ed., 2008).

92. LORI B. ANDREWS, MAXWELL J. MEHLMAN & MARK A. ROTHSTEIN, GENETICS: ETHICS, LAW, AND POLICY 301 (4th ed. 2015); *What Are the Types of Genetic Tests?*, U.S. NAT'L LIBR. MED. (Aug. 6, 2019), <https://ghr.nlm.nih.gov/primer/testing/uses> [<https://perma.cc/APM9-3U8T>].

93. A mutation "is a permanent alteration in the DNA sequence that makes up a gene, such that the sequence differs from what is found in most people." *What Is a Gene Mutation and How Do Mutations Occur?*, U.S. NAT'L LIBR. MED. (Aug. 6, 2019), <https://ghr.nlm.nih.gov/primer/mutationsanddisorders/genemutation> [<https://perma.cc/C536-HMW4>].

94. See Press, *supra* note 91, at 73; Lydia Ramsey, *Over a 40-Year Career, This 'Stubborn Scientist' Helped Change the Way We Think About Cancer and Genetics*, BUS. INSIDER (Nov. 16, 2017), <https://www.businessinsider.com/mary-claire-king-and-the-impact-brca-genes-had-on-cancer-genetics-2017-11> [<https://perma.cc/DC2G-Z7GL>]; *BRCA1 & BRCA2 Genes: Risk for Breast & Ovarian Cancer*, MEMORIAL SLOAN KETTERING CANCER CTR., <https://www.mskcc.org/cancer-care/risk-assessment-screening/hereditary-genetics/genetic-counseling/brca1-brca2-genes-risk-breast-ovarian> [<https://perma.cc/ZC3R-Z6YK>].

95. See ANDREWS ET AL., *supra* note 92, at 301; Burke, *supra* note 88, at 1867, 1869–70.

variety of cancers, hereditary hemochromatosis (a disorder causing iron overload), Huntington's disease, and more.⁹⁶

B. Genetic Testing Concerns

The advent of genetic testing raised numerous concerns that were vigorously debated and that catapulted professional and governmental bodies into action. Academics wrote hundreds of articles about genetic testing, and law reviews dedicated entire symposium issues to the subject.⁹⁷ In 1995, the National Institutes of Health-Department of Energy Joint Working Group on the Ethical, Legal and Social Implications of Human Genome Research established the Task Force on Genetic Testing.⁹⁸ The American Academy of Ophthalmology convened such a task force as well.⁹⁹ For purposes of illustration, this section will focus on three of the many concerns that were considered: clinical validity and accuracy; privacy and discrimination; and psychological harms.

1. Clinical Validity and Accuracy

Experts worry about the clinical validity and accuracy of genetic test results.¹⁰⁰ As discussed in Section III.C below, predictive health analytics raises similar concerns. Many genetic tests identify only a fraction of genetic mutations that can cause a disease because researchers have yet to discover other mutations or because the price of more comprehensive testing is too high.¹⁰¹ Moreover, although a subgroup of patients may have an inherited form of a disease such as cancer, many others will develop the disease because of

96. See Burke, *supra* note 88, at 1870; *About Huntington's Disease*, NAT'L HUM. GENOME RES. INST. (Nov. 17, 2011), <https://www.genome.gov/10001215/learning-about-huntingtons-disease/> [<https://perma.cc/673J-QCAC>]; *What Are the Types of Genetic Testing?*, *supra* note 92.

97. See, e.g., *The Fifth Annual Health Law Symposium Communities of Color and Genetic Testing: Purpose, Voice, & Values*, 27 SETON HALL L. REV. 887 (1997) (discussing genetic testing and its possible ramifications); Symposium, *The Genetics Revolution: Conflicts, Challenges and Conundra*, 28 AM. J.L. & MED. 145 (2002); Symposium, *Legal and Ethical Issues Raised by the Human Genome Project*, 29 HOUS. L. REV. 1 (1992); Symposium, *Legal Liabilities at the Frontier of Genetic Testing* (pts. I & II), 41 JURIMETRICS 1 (2000), 41 JURIMETRICS 145 (2001); Symposium, *Living in the Genetic Age: New Issues, New Challenges*, 3 ST. LOUIS U. J. HEALTH L. & POL'Y 1 (2009).

98. *Task Force on Genetic Testing*, NAT'L HUM. GENOME RES. INST. (Apr. 1995), <https://www.genome.gov/10001808/genetic-testing-task-force/> [<https://perma.cc/3VRR-G3XG>].

99. See Edwin M. Stone et al., *Recommendations for Genetic Testing of Inherited Eye Diseases*, 119 OPTHALMOLOGY 2408, 2408 (2012) (providing a report drafted by the task force convened by the American Academy of Ophthalmology).

100. See Burke, *supra* note 88, at 1871; Neil A. Holtzman, *Promoting Safe and Effective Genetic Tests in the United States: Work of the Task Force on Genetic Testing*, 45 CLINICAL CHEMISTRY F. 732, 737 (1999).

101. See Burke, *supra* note 88, at 1871.

environmental or other triggers without having genetic mutations.¹⁰² Individuals who undergo genetic testing and receive negative results may mistakenly conclude that they are immune to the disease at issue. Thus, a woman who is found not to have the BRCA1 or BRCA2 mutation may decline appropriate screening measures, such as routine mammograms and gynecological exams believing she is not at risk. In truth, however, only five to ten percent of breast and ovarian cancers are hereditary.¹⁰³

Another risk is that the opposite will occur. An individual who receives a positive genetic test result may panic and take unnecessarily aggressive preventive measures.¹⁰⁴ Many genetic mutations are not completely penetrant; that is, not all individuals with the abnormality will develop the disease at issue.¹⁰⁵ For example, a woman who tests positive for the BRCA1 mutation has only a fifty-five to sixty-five percent chance of developing breast cancer by the age of seventy.¹⁰⁶ Women who fully understand the meaning of their test results and the extent of their risk may or may not want to undergo prophylactic radical mastectomies, and either decision would be rational.

Physicians and patients who use genetic testing must be fully educated about how to interpret test results and the limitations of the information they reveal.¹⁰⁷ It is all too easy to misconstrue test outcomes and attribute more certainty to genetic predictions than they warrant.¹⁰⁸ Such misunderstandings can lead to consequential medical treatment missteps.

2. Privacy and Discrimination

The dearth of regulation designed to protect patients against medical privacy violations and genetic discrimination led to significant concern in legal and policy circles for several decades.¹⁰⁹ Until 2003, there was no federal law

102. *BRCA1 & BRCA2 Genes: Risk for Breast & Ovarian Cancer*, *supra* note 94.

103. *Id.*

104. ANDREWS ET AL., *supra* note 92, at 301–02.

105. *What Are Reduced Penetrance and Variable Expressivity?*, U.S. NAT'L LIBR. MED. (Aug. 6, 2019), <https://ghr.nlm.nih.gov/primer/inheritance/penetranceexpressivity> [<https://perma.cc/DE7X-FYKZ>] (“Penetrance refers to the proportion of people with a particular genetic change (such as a mutation in a specific gene) who exhibit signs and symptoms of a genetic disorder.”).

106. *BRCA1 and BRCA2*, SUSAN G. KOMEN (Dec. 7, 2018), <https://ww5.komen.org/BreastCancer/BRCA1andBRCA2.html> [<https://perma.cc/R4XJ-FT2C>].

107. *See* Burke, *supra* note 88, at 1871 (discussing limitations).

108. *See id.*

109. *See generally* Eric Mills Holmes, *Solving the Insurance/Genetic Fair/Unfair Discrimination Dilemma in Light of the Human Genome Project*, 85 KY. L.J. 503, 507–08, 513–14, 578 (1997) (arguing that genetic mapping and testing could become a focal civil rights issue of the twenty-first century, potentially prompting the need for a legislative solution); Pauline T. Kim, *Genetic Discrimination, Genetic Privacy: Rethinking Employee Protections for a Brave New Workplace*, 96 NW. U. L. REV. 1497,

that safeguarded the privacy of health information in general, let alone genetic information in particular.¹¹⁰ Thus, federal law did not prohibit anyone who possessed genetic information from disclosing it to third parties.¹¹¹ At the state level, only a patchwork of statutes offered varying degrees of genetic privacy protections in some states.¹¹² Moreover, until the passage of the Genetic Information Nondiscrimination Act (“GINA”) in 2008,¹¹³ no federal law prohibited third parties, such as employers and health insurers, from demanding that individuals provide genetic information or from discriminating on its basis.¹¹⁴ While some states took the lead and passed genetic discrimination legislation as early as the 1970s, the protections they offered were inconsistent and often limited.¹¹⁵

Without comprehensive privacy protection, sensitive genetic information could end up in the hands of third parties that could use it to advance their own interests to the detriment of data subjects. The prospect of genetic discrimination generated a plethora of literature and many heated academic and policy debates.¹¹⁶

For example, workers worried that employers would obtain genetic data through pre- or post-employment medical examinations.¹¹⁷ Workers were

1498, 1523 (2002) (addressing concerns about genetic discrimination in employment and explaining that carriers of the sickle cell trait often suffered discrimination because of a lack of privacy protections).

110. HOFFMAN, *supra* note 29, at 62 (discussing the HIPAA Privacy Rule, which became effective in 2003); Karen H. Rothenberg, *Breast Cancer, the Genetic “Quick Fix,” and the Jewish Community: Ethical, Legal, and Social Challenges*, 7 HEALTH MATRIX 97, 115 (1997) (stating that at the time there was no federal law regarding health information).

111. See HOFFMAN, *supra* note 29, at 62 (stating that, after 2003, there were certain legislative provisions that regulated the dissemination of medical information to third parties); Rothenberg, *supra* note 110, at 115 (clarifying that there was no existing federal law in 1997 that addressed genetic privacy and confidentiality).

112. See Lori B. Andrews, *A Conceptual Framework for Genetic Policy: Comparing the Medical, Public Health, and Fundamental Rights Models*, 79 WASH. U. L.Q. 221, 280 (2001); Rothenberg, *supra* note 110, at 115–17.

113. Genetic Information Nondiscrimination Act of 2008, Pub. L. No. 110-233, 122 Stat. 881 (codified as amended in scattered sections of 29 and 42 U.S.C.).

114. Rothenberg, *supra* note 110, at 107–14.

115. *Id.* at 108–09, 114–15; see also Office of Biological and Env'tl. Research, U.S. Dep't of Energy Office of Sci., *Genetics Legislation*, HUM. GENOME PROJECT INFO. ARCHIVE 1990–2003, https://web.ornl.gov/sci/techresources/Human_Genome/elsi/legislat.shtml [<https://perma.cc/T4N9-AK69>] (last modified Apr. 23, 2019) (describing state legislation as a “patchwork” and explaining that some state laws “prohibited discrimination against individuals with specific genetic traits or disorders” while others “regulate both the use of genetic testing in employment decisions and the disclosure of genetic test results”).

116. See *supra* notes 97–100 and accompanying text.

117. Ellen R. Peirce, *The Regulation of Genetic Testing in the Workplace – A Legislative Proposal*, 46 OHIO ST. L.J. 771, 801–04 (1985).

concerned that, once they learned of individuals' genetic abnormalities, employers could reject, fire, demote, or otherwise discriminate against them with impunity.¹¹⁸

Americans were also apprehensive about the impact of genetic testing on health insurance coverage.¹¹⁹ An individual policy insurer who obtained data about an applicant's disease risks could potentially decline to insure the person, raise premium prices, or dictate other adverse coverage conditions.¹²⁰ The same could be true for other types of insurance, such as long-term care policies.¹²¹

3. Psychological Harms

A third area of concern centered on psychological harms. Individuals who discover they are at risk of a life-threatening disease may suffer depression and even become suicidal.¹²² They may lose motivation to be productive in their careers, experience diminished self-esteem, and have difficulty caring for their families.¹²³ Some may even decide not to get married or have children because they expect to die young and do not wish to transmit a genetic abnormality to a child.¹²⁴ On the other hand, for patients who build their lives around the assumption that they will inherit a disease that runs in their family, such as breast cancer or Huntington's disease,¹²⁵ obtaining a negative genetic test result

118. *See id.* While these concerns were prevalent, there was little evidence that genetic discrimination was actually taking place. *See* Jessica L. Roberts, *The Genetic Information Nondiscrimination Act as an Antidiscrimination Law*, 86 NOTRE DAME L. REV. 597, 625 (2011) (explaining that "instead of reacting to current discrimination like its predecessors, GINA is a forward-looking statute—designed to preempt a variety of discrimination before it becomes entrenched").

119. Andrews, *supra* note 112, at 258–60. This concern was prevalent because unlike other developed countries, the United States does not have universal health coverage. *See Land of the Free-for-All: America is a Health-Care Outlier in the Developed World*, ECONOMIST (Apr. 26, 2018), <https://www.economist.com/special-report/2018/04/26/america-is-a-health-care-outlier-in-the-developed-world> [<https://perma.cc/37QG-QJA6> (dark archive)].

120. *See Andrews, supra* note 112 at 280 (explaining that at the time the law did "not prohibit genetic discrimination against people seeking insurance under individual plans" and did "not prohibit group insurers from charging higher rates to a whole group based on genetic information about a particular individual"); Robert Lowe, *Genetic Testing and Insurance: Apocalypse Now?*, 40 DRAKE L. REV. 507, 510–11 (1991).

121. *See* Mark A. Rothstein, *Predictive Genetic Testing for Alzheimer's Disease in Long-Term Care Insurance*, 35 GA. L. REV. 707, 707–08 (2001).

122. *See* Kathryn M. Kash, *Psychosocial and Ethical Implications of Defining Genetic Risk for Cancers*, 768 ANNALS N.Y. ACAD. SCI. 41, 45–46 (1995) (discussing "psychological issues in women at genetic risk"); Katherine A. Schneider, *Adverse Impact of Predisposition Testing on Major Life Activities: Lessons from BRC1/2 Testing*, 3 J. HEALTH CARE L. & POL'Y 365, 369, 372–74 (2000).

123. *See* Schneider, *supra* note 122, at 369, 372–75.

124. *Id.* at 376.

125. *See id.* at 374.

may be equally devastating.¹²⁶ They may be confused and depressed by the need to reorient their lives and feel “survivor guilt” in the face of their loved ones’ suffering.¹²⁷

The risk of psychological injury is particularly acute for minors, especially when tested for adult-onset illnesses, such as Huntington’s disease.¹²⁸ Experts questioned whether it was ethical to test individuals under the age of eighteen with or without the minor’s assent.¹²⁹ They also pondered who should gain access to test results and the extent to which clinicians should ask both parents and their children to consent to the testing.¹³⁰

If preventive measures such as regular screening and curative medical interventions are available, genetic testing of children can be justified and beneficial.¹³¹ However, in the absence of such measures, the American Society of Human Genetics recommended against testing because of potential negative psychosocial implications for minors and their family members.¹³² Knowing that they live in the shadow of an impending illness could ravage minors’ psychological well-being.¹³³ Likewise, discovery of a child’s genetic abnormality may upend family dynamics as the affected child may be treated either as more precious than others or less favorably because the child does not have a promising future.¹³⁴ Admittedly, however, in the absence of genetic testing, families in which a genetic disease is prevalent may make assumptions about a child’s predisposition to the illness and experience some of the same stress and inequities.¹³⁵

126. *See id.*

127. *See id.*

128. Kimberly A. Quaid, *Genetic Testing for Huntington Disease*, in 144 HANDBOOK OF CLINICAL NEUROLOGY 113, 114–17 (A.S. Feigin & K.E. Anderson eds., 2017). For information about Huntington’s disease, see *Huntington’s Disease*, MAYO CLINIC (May 16, 2018), <https://www.mayoclinic.org/diseases-conditions/huntingtons-disease/symptoms-causes/syc-20356117> [https://perma.cc/B62Q-2HSG].

129. ANDREWS ET AL., *supra* note 92, at 331–32.

130. *Id.* at 331.

131. See Jeffrey R. Botkin et al., *Points to Consider: Ethical, Legal, and Psychosocial Implications of Genetic Testing in Children and Adolescents*, 97 AM. J. HUM. GENETICS 6, 7 (2015).

132. *See id.* at 8.

133. *See id.*

134. See Am. Soc’y of Human Genetics Bd. of Dirs. & Am. Coll. of Med. Genetics Bd. of Dirs., *Points to Consider: Ethical, Legal, and Psychosocial Implications of Genetic Testing in Children and Adolescents*, 57 AM. J. HUM. GENETICS 1233, 1236 (1995). At the same time, minors whose test results are negative “may develop ‘survivor guilt,’ based on the knowledge that one or more of their siblings will develop—and perhaps die from—a serious genetic disease.” *Id.*

135. See Botkin et al., *supra* note 131, at 7.

C. *Legal and Policy Interventions*

As genetic testing became increasingly common, legislators and other policymakers implemented a variety of measures to address the concerns that it raised.¹³⁶ This part will focus on three of these measures: state and federal antidiscrimination legislation, the HIPAA Privacy Rule, and self-regulation mechanisms.

1. State and Federal Antidiscrimination Legislation

States began enacting legislation to prohibit genetic discrimination as early as the 1970s.¹³⁷ Early laws focused on protecting individuals with the sickle cell trait.¹³⁸ In 1991, Wisconsin was the first state to enact a more comprehensive statute that prohibited employers from requiring any genetic testing or discriminating against workers who undergo genetic tests.¹³⁹ Thereafter, the vast majority of states enacted genetic antidiscrimination statutes, though they varied significantly in scope and content.¹⁴⁰ As applied to health insurers, these laws imposed restrictions on using genetic information to determine coverage eligibility or premium levels, requiring applicants to undergo genetic testing, or disclosing genetic information to others without consent.¹⁴¹ As applied to employers, the laws prohibited employers from discriminating on the basis of genetic information and from requesting, requiring, or obtaining genetic information.¹⁴²

Congress considered genetic discrimination bills for thirteen long years.¹⁴³ Finally, President George W. Bush signed GINA into law on May 21, 2008.¹⁴⁴ GINA applies to the use of predictive genetic information by health insurers

136. Office of Biological and Env'tl. Research, U.S. Dep't of Energy Office of Sci., *supra* note 115.

137. ANDREWS ET AL., *supra* note 92, at 776.

138. *See id.*

139. *Id.*; see WIS. STAT. ANN. § 111.372 (West, Westlaw through 2019 Act 5).

140. *See Genetics and Health Insurance State Anti-Discrimination Laws*, NAT'L CONF. ST. LEGISLATURES (Jan. 2008), <http://www.ncsl.org/research/health/genetic-nondiscrimination-in-health-insurance-laws.aspx> [<https://perma.cc/5YU2-47Z8>] [hereinafter *Genetics and Health Insurance*]; *see also Genetic Employment Laws*, NAT'L CONF. ST. LEGISLATURES (Jan. 2008), <http://www.ncsl.org/research/health/genetic-employment-laws.aspx> [<https://perma.cc/VN6J-TGKC>].

141. *See Genetics and Health Insurance*, *supra* note 140.

142. *See Genetic Employment Laws*, *supra* note 140.

143. ANDREWS ET AL., *supra* note 92, at 777.

144. Genetic Information Nondiscrimination Act of 2008, Pub. L. No. 110-233, 122 Stat. 881 (codified as amended in scattered sections of 29 and 42 U.S.C.); *President Bush Signs the Genetic Information Nondiscrimination Act of 2008*, NAT'L HUM. GENOME RES. INST. (May 21, 2008), <https://www.genome.gov/27026050/president-bush-signs-the-genetic-information-nondiscrimination-act-of-2008/> [<https://perma.cc/Q4ZR-RJN5>].

and employers.¹⁴⁵ The law does not cover those who already manifest symptoms of a genetic disease.¹⁴⁶

Title I of the Act prohibits genetic discrimination in health insurance.¹⁴⁷ Health insurers offering group plans may not modify premium prices and contribution amounts based on genetic information.¹⁴⁸ Similarly, insurers offering individual health plans may not require genetic testing or use genetic information to establish rules for eligibility, premium prices, or contribution amounts, or to apply preexisting condition exclusions for coverage.¹⁴⁹

GINA's Title II focuses on employment discrimination,¹⁵⁰ prohibiting employers from discriminating against employees in hiring, firing, or other employment practices based on genetic information.¹⁵¹ The law defines "genetic information" as including genetic testing of both individuals and their family members, as well as family disease histories.¹⁵² Furthermore, Title II prohibits employers from attempting to obtain genetic information about applicants or employees by requesting, requiring, or purchasing it.¹⁵³

GINA has many critics who decry its arguably anemic protections.¹⁵⁴ For example, it applies only to health insurers and employers rather than to all parties that might possess genetic information (such as life or disability insurers) and might subject individuals to discrimination on its basis.¹⁵⁵ GINA

145. Genetic Information Nondiscrimination Act §§ 101–04, 201–03, 122 Stat. at 833–903, 905–10 (codified as amended in scattered sections of 26, 29, and 42 U.S.C.); *President Bush Signs the Genetic Information Nondiscrimination Act of 2008*, *supra* note 144.

146. Mark A. Rothstein, *GINA's Beauty Is Only Skin Deep*, 22 GENE WATCH 9, 10 (2009) ("The problem is that GINA only applies to asymptomatic individuals."). Many symptomatic individuals, however, will be protected by the Americans with Disabilities Act ("ADA"). 42 U.S.C. § 12102(1) (2012) (defining "disability" under the ADA).

147. 29 U.S.C. § 1182 (2012).

148. *Id.* § 1182(b)(3).

149. 42 U.S.C. § 300gg-53(a)–(d)(1) (2012).

150. *Id.* § 2000ff-1.

151. *Id.* § 2000ff-1(a).

152. *Id.* § 2000ff(4)(A) (defining genetic information as "(i) [an] individual's genetic tests, (ii) the genetic tests of family members of [an] individual, and (iii) the manifestation of a disease or disorder in family members of [an] individual").

153. *Id.* § 2000ff-1(b).

154. *See, e.g.*, Bradley A. Areheart & Jessica L. Roberts, *GINA, Big Data, and the Future of Employee Privacy*, 128 YALE L.J. 710, 745 (2019) (noting that "the scholarly reaction to GINA has been almost entirely negative"); Russell Korobkin & Rahul Rajkumar, *The Genetic Information Nondiscrimination Act — A Half-Step Toward Risk Sharing*, 359 NEW ENG. J. MED. 335, 337 (2008) (criticizing "[t]he arbitrary nature of the categories GINA creates"); Rothstein, *supra* note 146, at 9 ("Unfortunately, the protections afforded individuals under either state laws prohibiting genetic discrimination in health insurance or GINA are not particularly robust or valuable.").

155. *See* 29 U.S.C. § 1182(b)(3) (2012); 42 U.S.C. §§ 300gg-53(a)–(d)(1), 2000f(4)(A), 2000ff-1(a)–(b).

is also unlikely to cover a range of non-genetic biologic information that may be of interest to third parties, such as epigenetic markers and the microbiome.¹⁵⁶

A full analysis of GINA or parallel state legislation is beyond the scope of this Article. For the purposes of this Article, GINA is relevant only to demonstrate that legislators recognized that genetic testing could yield both benefits and serious risks. They were sufficiently thoughtful and concerned about those risks to enact statutory interventions, however imperfect. Policymakers should be equally responsible in tackling the risks of long-term predictive health analytics now.¹⁵⁷

2. The HIPAA Privacy Rule

The HIPAA Privacy Rule, which went into effect in 2003, is a set of federal regulations that addresses the privacy of health information.¹⁵⁸ The Privacy Rule establishes that, with some exceptions, “covered entities”¹⁵⁹ must obtain patients’ permission to disclose their protected health information to third parties.¹⁶⁰ As of 2013, “health information” explicitly includes genetic information.¹⁶¹

Under the HIPAA Privacy Rule, covered entities must allow patients to view and obtain copies of their health records and receive an accounting of disclosures of their protected health information.¹⁶² In addition, patients can ask healthcare providers to correct errors in their medical records or to use their health data restrictively.¹⁶³ Covered entities that suffer privacy breaches of

156. Areheart & Roberts, *supra* note 154, at 748–49 (stating that “at present it is unclear whether GINA covers epigenetic markers, the microbiome, or myriad other kinds of biological information related to new technologies”).

157. See *infra* Parts III & IV (discussing predictive health analytics concerns and suggested legislative interventions).

158. 45 C.F.R. §§ 160.101–.534 (2018); *HIPAA for Professionals*, U.S. DEP’T HEALTH & HUM. SERVS. (June 16, 2017), <https://www.hhs.gov/hipaa/for-professionals/index.html> [<https://perma.cc/T2WS-FTS5>]. The HIPAA Privacy Rule was promulgated pursuant to the Health Insurance Portability and Accountability Act of 1996 and was amended in accordance with the Health Information Technology for Economic and Clinical Health (“HITECH”) Act of 2009; see Health Information Technology for Economic and Clinical Health Act, Pub. L. No. 111-5, div. A, tit. XIII & div. B, tit. IV, 123 Stat. 226, 467 (2009) (codified as amended in scattered sections of 42 U.S.C.); see also American Recovery and Reinvestment Act of 2009 (“ARRA”), Pub. L. No. 111-5, 123 Stat. 226 (codified as amended in scattered sections of 42 U.S.C.).

159. For a definition of covered entities, see 42 U.S.C. § 17934 (2012); 45 C.F.R. §§ 160.102–.103. See also HOFFMAN, *supra* note 29, at 73.

160. See 45 C.F.R. §§ 164.508–.510 (2018).

161. *Id.* § 160.103; *Genetic Information Privacy*, ELECTRONIC FRONTIER FOUND., <https://www.eff.org/issues/genetic-information-privacy> [<https://perma.cc/TP3E-ZDPH>].

162. 45 C.F.R. §§ 164.520(b)(1)(iv), .528 (2018).

163. *Id.* §§ 164.520(b)(1)(iv), .522(a)(1).

unsecured data, such as hacking occurrences, must notify affected individuals, the Department of Health and Human Services, and, in instances of large breaches, the media.¹⁶⁴

The related HIPAA Security Rule, which became effective in 2005, promotes secure storage and processing of electronic health information (“EHI”).¹⁶⁵ It delineates administrative, physical, and technical safeguards to protect EHI’s confidentiality, integrity, and availability.¹⁶⁶

The HIPAA Privacy and Security Rules offer valuable protections to American patients. However, like genetic antidiscrimination statutes, they are limited in scope and have been subject to criticism.¹⁶⁷ For example, “covered entity” includes only “a health plan, . . . a health care clearinghouse, . . . a health care provider who transmits any health information” electronically for purposes of HIPAA-relevant transactions, and their business associates.¹⁶⁸ Other parties that possess and handle health information, such as data brokers and marketers, need not comply with the Rules’ privacy and security mandates.¹⁶⁹

Another noteworthy regulatory gap in the HIPAA Privacy and Security Rules is the absence of a private cause of action.¹⁷⁰ Thus, individuals whose health data is breached cannot sue wrongdoers for damages under federal law no matter what consequences they suffer.¹⁷¹ Instead, the regulations leave enforcement solely in the hands of the Department of Health and Human Services: Office for Civil Rights and state attorneys general offices,¹⁷² which may or may not have adequate staffing and resources for robust prosecutorial activities.

In addition, the HIPAA Privacy Rule features numerous exceptions. Covered entities can disclose patients’ medical data for purposes of treatment, payment, and healthcare operations without patient authorization.¹⁷³ Thus, the regulations permit physicians to consult colleagues about patients and to ask

164. *Id.* §§ 164.400–408. Entities must notify the media if a breach involves “more than 500 residents of a State or jurisdiction.” *Id.* § 164.408(a). “Unsecured protected health information” means information “that is not rendered unusable, unreadable, or indecipherable to unauthorized persons through the use of a [specified] technology or methodology[,]” such as encryption. *Id.* § 164.402.

165. Sharona Hoffman & Andy Podgurski, *In Sickness, Health, and Cyberspace: Protecting the Security of Electronic Private Health Information*, 48 B.C. L. REV. 331, 335–36 (2007).

166. 45 C.F.R. §§ 164.302–.318 (2018).

167. *See, e.g.*, Hoffman & Podgurski, *supra* note 165, at 344–59.

168. 45 C.F.R. § 160.103 (2018). *See also* 42 U.S.C. § 17934 (2012); 45 C.F.R. § 160.102 (2018).

169. *See* HOFFMAN, *supra* note 29, at 73.

170. *See id.* at 75.

171. *Id.*

172. 42 U.S.C. § 1320d-5(d) (2012); 45 C.F.R. §§ 160.300, .306.

173. 45 C.F.R. § 164.506(a) (2018).

administrators to review records for billing or other office-related purposes without the patients' knowledge.¹⁷⁴

The rule exempts additional disclosures as well, such as those made for law enforcement, public health, and other listed purposes.¹⁷⁵ In general, these exceptions are reasonable and sound. However, patients should understand that they are often unaware of who is viewing their health data and for what purpose.

The HIPAA Privacy and Security Rules fall short of providing patients with comprehensive protection. Nevertheless, they constitute important advances in the privacy arena and address some of the concerns raised by genetic testing. The HIPAA Rules' application to predictive health analytics and their gaps in this context will be analyzed in Section III.B below.

3. Self-Regulation

Self-regulation by genetic testing professionals provides another source of genetic testing constraints. Professional organizations have authored practice guidelines, and providers routinely defer testing until they have educated patients about its potential consequences. For example, the American Society of Breast Surgeons issued guidance that formulated recommendations for genetic testing related to breast cancer and discussed testing limitations.¹⁷⁶ Likewise, the American Society of Human Genetics Board of Directors and the American College of Medical Genetics Board of Directors published a document that discourages testing children for adult-onset diseases if they will derive no medical or psychological benefit from being tested as minors.¹⁷⁷ The American College of Medical Genetics and Genomics issued practice guidelines related to genetic testing for numerous conditions.¹⁷⁸ Unfortunately, medical practices do not always implement clinical practice guidelines effectively or

174. *Id.* § 164.506(c).

175. *Id.* §§ 164.502, .512.

176. *Consensus Guideline on Genetic Testing for Hereditary Breast Cancer*, AM. SOC'Y BREAST SURGEONS 1–2, 5 (2019), https://www.breastsurgeons.org/about/statements/PDF_Statements/Hereditary_Genetic_Testing_Patients_With_Without_Breast_Cancer.pdf [<https://perma.cc/26SV-G36U>].

177. Botkin et al., *supra* note 131, at 7; *see also* Quaid, *supra* note 128, at 115–17 (discussing testing guidelines for Huntington's disease, including those addressing predictive testing of minors).

178. *Practice Guidelines*, AM. C. MED. GENETICS & GENOMICS (2019), <https://www.acmg.net/ACMG/Medical-Genetics-Practice-Resources/Practice-Guidelines.aspx> [<https://perma.cc/ARB5-4LFF> (staff-uploaded archive)].

consistently.¹⁷⁹ However, when healthcare providers adhere to sound clinical practice guidelines, they can often achieve improved treatment outcomes.¹⁸⁰

Clinicians typically offer counseling to patients who are considering genetic testing.¹⁸¹ Counseling is designed to ensure that patients make fully informed decisions about pursuing testing in light of the benefits and risks that exist in their particular circumstances.¹⁸² A variety of healthcare providers can educate patients about genetic testing, but a growing number of practices include professional genetic counselors with master's degrees.¹⁸³ The American Board of Genetic Counselors has certified over 4000 genetic counselors thus far.¹⁸⁴ Therefore, rather than rushing to test patients after only a brief discussion, responsible clinicians exercise a degree of self-restraint and take the steps necessary to ensure that patients provide meaningful and genuinely informed consent to the procedure.

III. LONG-TERM PREDICTIVE HEALTH ANALYTICS CONCERNS

Like genetic testing, long-term predictive health analytics is fraught with risks but, unlike the perils of genetic testing, these are garnering too little attention.¹⁸⁵ This part highlights three areas of concern: psychological harms, privacy and discrimination, and erroneous predictions. It concludes by analyzing the U.S. Food and Drug Administration's ("FDA") regulatory power over long-term predictive health analytics.

179. Fargol Mostofian et al., *Changing Physician Behavior: What Works?*, 21 AM J. MANAGED CARE 75, 82 (2015) ("Various implementation methods are utilized to try to change physician behavior, and implementing the most effective ones is crucial to success."); M. Hassan Murad, *Clinical Practice Guidelines: A Primer on Development and Dissemination*, 92 MAYO CLINIC PROC. 423, 423, 429–31 (2017) ("[G]uidelines require active dissemination and innovative implementation strategies.").

180. William J. Hanney et al., *The Influence of Physical Therapy Guideline Adherence on Healthcare Utilization and Costs Among Patients with Low Back Pain: A Systematic Review of the Literature*, 11 PLOS ONE e0156799, 1–2, 15 (2016); Murad, *supra* note 179, at 429; Jannicke Sletli Wathne et al., *The Association Between Adherence to National Antibiotic Guidelines and Mortality, Readmission and Length of Stay in Hospital Inpatients: Results from a Norwegian Multicentre, Observational Cohort Study*, 8 ANTIMICROBIAL RESISTANCE & INFECTION CONTROL 1, 4, 8 (2019).

181. See Burke, *supra* note 88, at 1873.

182. *Id.*

183. See ANDREWS ET AL., *supra* note 92, at 329; *About NSGC: The Leading Voice for Genetic Counselors*, NAT'L SOC'Y GENETIC COUNS., <https://www.nsgc.org/page/about-nsgc> [<https://perma.cc/4RZR-A9WH>].

184. *Mission Statement, Purpose and Values*, AM. BOARD GENETIC COUNS., <https://www.abgc.net/about-abgc/mission-history/> [<https://perma.cc/F2RQ-5TSK>].

185. See I. Glenn Cohen & Harry S. Graver, *Cops, Docs, and Code: A Dialogue Between Big Data in Health Care and Predictive Policing*, 51 U.C. DAVIS L. REV. 437, 446 (2017) ("The legal literature on predictive analytics in health care is at this moment less robust than that on predictive policing, although that is changing.").

A. *Psychological Harms*

Predictions of future ailments based on predictive health analytics can be just as traumatizing as predictions based on genetic testing.¹⁸⁶ Individuals who learn from their doctors that they are likely to develop heart disease, dementia, or psychosis in the future might find the news devastating.¹⁸⁷ As a result, they could have difficulty concentrating on work, experience strain in their relationships, or even become clinically depressed or suicidal.¹⁸⁸ Like genetic testing of children, predictive health analytics involving minors raises particularly troubling questions.¹⁸⁹ Worrisome predictions can adversely impact children's futures and disrupt family dynamics.¹⁹⁰

Furthermore, physicians who identify certain individuals as vulnerable to opioid addiction, cognitive decline, or suicide¹⁹¹ may treat those patients differently, to the patients' detriment. For example, they may refuse to provide potential opioid addicts with needed pain medication.¹⁹² They may also relate poorly to patients at risk of dementia or suicide, treating them as cognitively compromised or lacking autonomy even when they are fully competent.¹⁹³ So too, clinicians may try to drive patients who are labelled as potentially high-risk and high-cost away from their practices.¹⁹⁴

It is also possible that individuals will obtain distressing health predictions not from their doctors but from commercial enterprises without being aware in advance that anyone has assessed their health risks.¹⁹⁵ Data brokers sell health information to interested buyers, and companies such as LexisNexis and Acxiom have already begun to engage in predictive health analytics.¹⁹⁶ Marketers will likely be eager to obtain health predictions about patients in order to tailor their marketing materials effectively.¹⁹⁷ Imagine individuals

186. See *supra* Section I.B.

187. See *supra* Section I.B.

188. See *supra* notes 122–26 and accompanying text.

189. See *supra* notes 128–34 and accompanying text.

190. See *supra* notes 128–34 and accompanying text.

191. See *supra* Section I.B.

192. See Ravindranath, *Risk Scores*, *supra* note 80.

193. Marks, *Artificial Intelligence*, *supra* note 13 (manuscript at 24) (“Patients with mental illnesses often report feeling dehumanized and dismissed by healthcare providers.”).

194. Cohen et al., *supra* note 6, at 1141 (“For instance, the data could be used to identify vulnerable high-risk, high-cost patients and exclude them from care.”).

195. See *supra* notes 80–87 and accompanying text.

196. See *supra* notes 80–87 and accompanying text.

197. See HOFFMAN, *supra* note 29, at 60; Cobb, *supra* note 87. Commercial enterprises already use data mining for marketing purposes. In one well-known case, Target sent a teenage girl advertisements for baby goods after determining that she was pregnant based on her prior purchases. See Kashmir Hill, *How Target Figured out a Teen Girl Was Pregnant Before Her Father Did*, FORBES (Feb. 16, 2012),

receiving the news that they are at risk of cognitive decline through an electronic advertisement urging them to purchase memory-enhancing products. People who do not have the support of a physician and do not receive a clear, medical explanation of the prediction and its degree of certainty will be all the more vulnerable to distress and misunderstandings.

B. *Privacy and Discrimination*

Because the HIPAA Privacy Rule governs only a subset of parties that possess health information,¹⁹⁸ not all predictive health analytics outcomes will be subject to privacy protections.¹⁹⁹ Entities that are not health plans, healthcare clearinghouses, healthcare providers, or their business associates are not legally bound to refrain from disclosing health information about patients.²⁰⁰ Thus, data brokers are permitted to sell health-related information to marketers.²⁰¹ Moreover, entities that are not covered by HIPAA could disclose and publicize individually identifiable predictive health analytics results. One can imagine the media obtaining predictions about entertainers and politicians that could cause significant embarrassment and even ruin careers. Predictions about ordinary people could likewise be widely publicized through social media and be available to anyone with a smart device.

It is also noteworthy that noncovered entities are not subject to the requirements of the HIPAA Security Rule.²⁰² As such entities may be tempted to use security shortcuts, they may be vulnerable to hacking and other data breaches.²⁰³ Consequently, data stored by commercial enterprises for predictive health analytics purposes may be more vulnerable to privacy violations than HIPAA-protected health information.

Given the HIPAA Privacy Rule's limitations, individuals' health predictions can easily land in the hands of third parties who may use them to

<https://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/#50c347c26668> [https://perma.cc/WL24-AVWZ]. The girl's father, who saw the mail but was unaware of his daughter's condition, angrily confronted a Target manager. *Id.* But, after later speaking with his daughter, he learned that Target was right. *Id.*

198. See *supra* note 168 and accompanying text.

199. See Marks, *Artificial Intelligence*, *supra* note 13 (manuscript at 6) (discussing privacy risks related to suicide predictions).

200. See *id.*; see also 42 U.S.C. § 17934 (2012); 45 C.F.R. §§ 160.102–.103 (2018).

201. See Erin McCann, *What HIPAA Doesn't Cover*, HEALTHCARE IT NEWS (July 16, 2014, 10:58 AM), <https://www.healthcareitnews.com/news/what-hipaa-doesnt-cover> [https://perma.cc/CK4G-7DLA].

202. Sharona Hoffman & Andy Podgurski, *supra* note 165, at 344–47.

203. See *id.* at 332–34 (discussing various data breaches); Topol, *supra* note 5, at 52 (noting “the risk of deliberate hacking of an algorithm to harm people at a large scale”).

further their own economic agendas.²⁰⁴ Employers, lenders, life insurers, and others with a stake in individuals' future well-being may be interested in predictions about individuals' health status in later years.²⁰⁵ Employers, for example, are interested in employees who will not have productivity or absenteeism problems and will not generate high health insurance costs.²⁰⁶ They may be very tempted to reject or terminate workers whom they believe to be at high risk of becoming seriously ill in the coming years. Similarly, lenders seek borrowers who will remain able to work and pay off their loans, and life insurers may use predictive information about applicants to make eligibility or pricing decisions.²⁰⁷

Currently, the antidiscrimination laws do not prohibit employers and others from discriminating based on predictions of future health problems (other than predictions based on genetic information, which are covered by GINA).²⁰⁸ The Americans with Disabilities Act ("ADA"), the primary federal disability discrimination law, prohibits discrimination related only to:

- (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 impairment²⁰⁹

Consequently, it does not reach discrimination based on future physical or mental impairments or disabilities. This legislative gap creates worrisome opportunities for discrimination based on disease predictions.

Predictive health analytics may also perpetuate other types of discrimination, such as sex- or race-based discrimination. Amazon's effort to develop artificial-intelligence-driven software to identify the best job candidates illustrates this point.²¹⁰ Because the predictive model's training data were past resumés submitted to Amazon mostly by men, the program was biased

204. HOFFMAN, *supra* note 29, at 59–60 (listing a variety of parties that could be interested in people's health data).

205. See 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, 85–86 (I. Glenn Cohen et al. eds., 2018) [hereinafter Hoffman, *New Discrimination*].

206. *Id.* at 86.

207. HOFFMAN, *supra* note 29, at 60; Marks, *Artificial Intelligence*, *supra* note 13 (manuscript at 11).

208. See Hoffman, *New Discrimination*, *supra* note 205, at 92–95.

209. 42 U.S.C. § 2102(1) (2012); Americans with Disabilities Act of 1990, Pub. L. No. 101-336, 104 Stat. 327 (codified as amended at 42 U.S.C. §§ 12101–12213 (2012)).

210. See Katherine Maher, *Without Humans, A.I. Can Wreak Havoc*, N.Y. TIMES (Mar. 12, 2019), <https://www.nytimes.com/2019/03/12/opinion/artificial-intelligence-wikipedia.html?smid=nytcore-ios-share> [https://perma.cc/8XN9-464W (dark archive)].

against women and concluded that men were preferable job candidates.²¹¹ It is likewise possible that predictive models in the healthcare arena will be biased and wrongly conclude that women are at higher risk of various health problems. Furthermore, if companies such as LexisNexis and Acxiom base predictive models on variables that include criminal records and voter registration data,²¹² they could disproportionately identify certain minorities as high-risk patients.²¹³ Thus, particular groups may be perceived as more prone to disease and biologically inferior to others.²¹⁴

C. *Erroneous Predictions*

As suggested above, the results of predictive health analytics can often be wrong,²¹⁵ just as genetic testing results can be misleading.²¹⁶ In one illustrative example outside the health field, scientists produced “fooling images” that were “completely unrecognizable to humans,” but deep neural networks (a form of machine learning) believed “with near certainty” that they were familiar objects.²¹⁷

211. *Id.* See *infra* Section III.C for a discussion of bias. Amazon acknowledged that its “recruiters looked at the recommendations generated by the tool when searching for new hires” but claimed they “never relied solely on those rankings” before Amazon abandoned the tool. Jeffrey Dastin, *Amazon Scraps Secret AI Recruiting Tool that Showed Bias Against Women*, REUTERS (Oct. 9, 2018), <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G> [https://perma.cc/WJ86-DUCF].

212. See Ravindranath, *Risk Scores*, *supra* note 80 (discussing risk scores based on other factors).

213. Minorities often have higher criminal conviction rates and lower voter turnout rates than whites. See THE SENT’G PROJECT, REPORT TO THE UNITED NATIONS ON RACIAL DISPARITIES IN THE U.S. CRIMINAL JUSTICE SYSTEM 1–4 (Apr. 19, 2018), <https://www.sentencingproject.org/publications/un-report-on-racial-disparities/> [https://perma.cc/3HDS-Q5QF] (discussing the racial disparity that exists in the U.S. criminal justice system, specifically for African Americans); Bernard L. Fraga, *The Turnout Gap Between Whites and Racial Minorities Is Larger Than You Think — and Hard to Change*, WASH. POST (Sept. 25, 2018), https://www.washingtonpost.com/news/monkey-cage/wp/2018/09/25/the-turnout-gap-between-whites-and-racial-minorities-is-larger-than-you-think-and-hard-to-change/?utm_term=.f46f666cc727 [https://perma.cc/8XN9-464W (dark archive)]. Algorithms that include these variables could consequently conclude that the relevant minorities are not as healthy as whites.

214. See Sharona Hoffman, “Racially-Tailored” Medicine Unraveled, 55 AM. U. L. REV. 395, 398 (2005) [hereinafter Hoffman, “Racially-Tailored”]; see also Jonathan Kahn, *How a Drug Becomes Ethnic: Law, Commerce, and the Production of Racial Categories in Medicine*, 4 YALE J. HEALTH POL’Y L. & ETHICS 1, 38 (2004) (“As race becomes correlated with various biological conditions, it takes only one further step to correlate race with a health threat.”).

215. See Ian A. Scott, *Hope, Hype and Harms of Big Data*, 49 INTERNAL MED. J. 126, 127 (2019) (noting ten potential harms or limitations of big data).

216. See *supra* Section II.B.1.

217. Anh Nguyen, Jason Yosinski & Jeff Clune, *Deep Neural Networks Are Easily Fooled: High Confidence Predictions for Unrecognizable Images*, in IEE CONFERENCE ON COMPUTER VISION &

Flawed outcomes can stem from a variety of problems. One cause can be human error in constructing or implementing the algorithm. Big data used to train computers and develop learning algorithms can be rife with inaccuracies and data gaps or otherwise be a poor fit for the task at hand.²¹⁸ Poor data quality will inevitably lead to poor data-driven artificial intelligence algorithms, consistent with the “garbage-in-garbage-out” principle.²¹⁹ For example, if the training data consist of electronic health records that are rife with errors, the algorithm will likely produce poor predictions.²²⁰

Moreover, learning algorithms can quickly become outdated.²²¹ As human knowledge advances or human behaviors change, analysts may need to update algorithms.²²² Outdated algorithms will not yield correct predictions.²²³ To illustrate, the emergence of vaping or a spike in the number of children who are not vaccinated might require modification of algorithms aimed at disease prediction.

Even with a correct learning algorithm, the predictive model’s performance when using the training data²²⁴ may not generalize to real world data because of a phenomenon called “overfitting.”²²⁵ A particular model can produce accurate predictions on a set of training data but fail to provide sound predictions when deployed on new data, especially if the model is complex and the training data set was small.²²⁶ Because of the dearth of training data and the large number of parameters used to construct the model, “the learned

PATTERN RECOGNITION 427, 427–28 (2015) (“Changing an image, originally correctly classified (e.g. as a lion), in a way imperceptible to human eyes, can cause a [Deep Neural Network] to label the image as something else entirely (e.g. mislabeling a lion as a library).”).

218. See Scott, *supra* note 215, at 127 (discussing numerous potential shortcomings of big data); Nilay D. Shah, Ewout W. Steyerberg & David M. Kemp, *Big Data and Predictive Analytics: Recalibrating Expectations*, 320 JAMA 27, 28 (2018); Topol, *supra* note 5, at 51.

219. Kun-Hsing Yu & Isaac S. Kohane, *Framing the Challenges of Artificial Intelligence in Medicine*, 28 BMJ QUALITY & SAFETY 238, 239 (2019); see also P. Elliott Miller et al., *Predictive Abilities of Machine Learning Techniques May Be Limited by Dataset Characteristics: Insights from the UNOS Database*, 25 J. CARDIAC FAILURE 479, 482 (2019) (“Our results raise the notion that large clinical datasets might lack the accuracy and granularity needed for machine learning methodologies to uncover unique associations.”); Rajkomar et al., *supra* note 34, at 1355.

220. See HOFFMAN, *supra* note 29, at 23–30 (discussing electronic health record shortcomings); see also Rajkomar et al., *supra* note 34, at 1355 (explaining that “[m]achine learning models generally perform better when they have access to large amounts of” error-free data); Yu & Kohane, *supra* note 219, at 239.

221. Sanjay P. Prabhu, *Ethical Challenges of Machine Learning and Deep Learning Algorithms*, 20 LANCET ONCOLOGY 621, 621 (2019).

222. *Id.*

223. *Id.*

224. See *supra* notes 34–37 and accompanying text for explanation of machine learning.

225. See CHARU C. AGGARWAL, NEURAL NETWORKS AND DEEP LEARNING 25 (2018).

226. See *id.*

parameters are spuriously inferred and are unlikely to *generalize* well” to unseen data.²²⁷ In other words, overfitting occurs when a predictive model fits the training data “too well.”²²⁸

Big data can also be subject to selection bias. If the data used to train learning algorithms or statistical models come from a health system that disproportionately serves particular populations (e.g., wealthy or disadvantaged individuals), the algorithm or model may not be generalizable to all patients.²²⁹ Several scholars have noted the following:

To date, Big Data has not captured certain marginalized demographics. Particularly concerning are racial minorities, people with low socioeconomic status, and immigrants. Many of the people missing from the data that come from sources such as Internet history, social media presence, and credit-card use are also missing from other sources of Big Data, such as electronic health records (EHRs) and genomic databases. The factors responsible for these gaps are diverse and include lack of insurance and the inability to access healthcare, to name just two²³⁰

It is even possible that attackers will hack into medical images and records and tamper with them.²³¹ Thus, malware could trick physicians into reaching incorrect conclusions about patients’ current or future illnesses.²³²

Unfortunately, it is often impossible to discern whether a predictive model is sound. Learning algorithms are often opaque because they rely on extremely complex rules and even their programmers are uncertain about how they ultimately work.²³³ Some commentators use the term “black box medicine” to describe reliance on nontransparent learning algorithms.²³⁴

Use of the terms “artificial intelligence” and “machine learning” can overawe people. But as one commentator notes, “[t]he only sure prediction

227. *Id.*

228. SHALEV-SHWARTZ & BEN-DAVID, *supra* note 22, at 16.

229. See Craig Konnoth, *Health Information Equity*, 165 U. PA. L. REV. 1317, 1361 (2017) (asserting that “relying on data that is biased towards certain social groups can have problematic effects”).

230. Sarah E. Malanga et al., *Who’s Left Out of Big Data? How Big Data Collection, Analysis, and Use Neglects Populations Most in Need of Medical and Public Health Research and Interventions*, in BIG DATA, HEALTH LAW, AND BIOETHICS, *supra* note 205, at 98, 98–99.

231. See Nancy Cohen, *Can Attackers Inject Malice into Medical Imagery? Fake Growths Here and There*, TECHXPLORE (Apr. 6, 2019), <https://techxplore.com/news/2019-04-malice-medical-imagery-fake-growths.html> [<https://perma.cc/2XSR-RCRW>].

232. *Id.*

233. Tokio Matsuzaki, *Ethical Issues of Artificial Intelligence in Medicine*, 55 CAL. W. L. REV. 255, 269 (2018) (“One concern is that AI [artificial intelligence] decision-making . . . often has no transparency. This means that doctors and patients are not able to know how the AI system reached the decision.”); Price, *supra* note 20, at 430.

234. See Price, *supra* note 20, at 429–30; Topol, *supra* note 5, at 51.

about the future of big data and predictive analytics is that it is unlikely to live up to some of the hype.²³⁵ Algorithms, in many cases, will falsely indicate that individuals may suffer particular conditions in the future, and the affected data subjects will be left to suffer the consequences.

D. *Regulatory Uncertainty*

Predictive health analytics may be particularly vulnerable to error because of deficient oversight. Genetic tests are subject to regulation by the FDA and the Centers for Medicare and Medicaid Services (“CMS”), which oversees clinical laboratories.²³⁶ CMS does not regulate learning algorithms because no clinical laboratories are involved,²³⁷ and a real question exists as to whether the FDA will routinely oversee predictive health analytics.

The FDA regulates medical devices, which are defined as any “instrument, apparatus, implement, machine, contrivance, implant, in vitro reagent, or other similar or related article . . . which is . . . intended for use in the diagnosis of disease or other conditions, or in the cure, mitigation, treatment, or prevention of disease, in man or other animals.”²³⁸ The FDA can also regulate “software as a medical device,” defined as “software intended to be used for one or more medical purposes that perform these purposes without being part of a hardware medical device.”²³⁹ For example, in early 2018, the FDA provided premarket clearance for the WAVE Clinical Platform, “an early-warning system” for hospitals, whose algorithm uses vital signs data to identify patients at risk of becoming unstable.²⁴⁰ The FDA also cleared Viz.AI’s Contact, which is software that uses an algorithm to analyze CT images for indications that patients are at

235. Shah et al., *supra* note 218, at 28.

236. *Regulation of Genetic Tests*, NAT’L HUM. GENOME RES. INST., <https://www.genome.gov/10002335/regulation-of-genetic-tests/> [<https://perma.cc/LH73-F3KU>].

237. See *Clinical Laboratory Improvement Amendments (CLIA)*, CTRS. FOR MEDICARE & MEDICAID SERVS., <https://www.cms.gov/Regulations-and-Guidance/Legislation/CLIA/index.html?redirect=/Clia> [<https://perma.cc/K55D-G2HJ>].

238. 21 U.S.C. § 321(h) (2012).

239. *Software as a Medical Device (SaMD)*, U.S. FOOD & DRUG ADMIN. (Aug. 31, 2018), <https://www.fda.gov/medical-devices/digital-health/software-medical-device-samd> [<https://perma.cc/E9HK-AYCK>].

240. Ravi B. Parikh, Ziad Obermeyer & Amol S. Navathe, *Regulation of Predictive Analytics in Medicine*, 363 SCI. 810, 811 (2019); *How WAVE Works*, EXCEL MED., <https://www.excel-medical.com/how-wave-works> [<https://perma.cc/Z7Y8-6XMR>].

risk of stroke.²⁴¹ In 2019, the FDA issued a proposal to improve its regulatory approach to algorithms that continuously learn and change over time.²⁴²

However, the agency is empowered to regulate only algorithms used for medical care.²⁴³ It would not have jurisdiction over predictive health analytics conducted by marketers, employers, or other parties for nonmedical purposes.²⁴⁴

Moreover, the FDA has traditionally refrained from regulating the practice of medicine.²⁴⁵ Thus, it may hesitate to regulate learning algorithms when their use seems akin to medical practice.²⁴⁶ While WAVE and the Viz.AI's Contact application may be classified as devices designed to predict imminent medical crises,²⁴⁷ long-term predictive analytics that help doctors anticipate illnesses that could develop later in life might be a poorer match. This is because long-term predictive analytics may not be perceived as close enough to traditional medical devices to warrant regulation.

Professor Nathan Cortez argues that predictive health analytics does not fit comfortably into any of the familiar categories of medical products, medical practice, or medical information for regulatory purposes.²⁴⁸ He and others argue for a new regulatory paradigm for predictive health analytics.²⁴⁹

Resolving the question of the extent to which the FDA will regulate long-term predictive health analytics is beyond the scope of this Article. Suffice it to say that there is uncertainty about the FDA's regulatory approach to such learning algorithms. Because long-term health predictions can significantly impact people's lives, they should not be ignored by regulators. The FDA

241. *FDA Clears First AI-Powered Clinical Decision Support Software for Stroke*, IMAGING TECH. NEWS (Feb. 14, 2018), <https://www.itnonline.com/content/fda-clears-first-ai-powered-clinical-decision-support-software-stroke> [<https://perma.cc/FH7Y-ZGYT>].

242. See U.S. FOOD & DRUG ADMIN., PROPOSED REGULATORY FRAMEWORK FOR MODIFICATIONS TO ARTIFICIAL INTELLIGENCE/MACHINE LEARNING (AI/ML)-BASED SOFTWARE AS A MEDICAL DEVICE (SAMD) - DISCUSSION PAPER AND REQUEST FOR FEEDBACK 2-4 (2019), <https://www.fda.gov/media/122535/download> [<https://perma.cc/R63X-4T59>]; Karen Hao, *The FDA Wants to Regulate Machine Learning in Health Care*, PREDICTIVE ANALYTICS TIMES (Apr. 6, 2019), <https://www.predictiveanalyticsworld.com/patimes/the-fda-wants-to-regulate-machine-learning-in-health-care/10268/> [<https://perma.cc/92SX-BW4Y>].

243. See *supra* note 238 and accompanying text.

244. See *supra* notes 85-87 and accompanying text.

245. Price, *supra* note 20, at 441.

246. *Id.* at 441-42; see also Cortez, *supra* note 30, at 78-79.

247. See Parikh et al., *supra* note 240, at 811.

248. Cortez, *supra* note 30, at 81; see also, Parikh et al., *supra* note 240, at 811 (stating that "existing FDA standards do not neatly translate to advanced predictive algorithms").

249. Cortez, *supra* note 30, at 81-85; Price, *supra* note 20, at 457-59 (calling for reform and suggesting that the FDA should regulate predictive health analytics in collaboration with other health care stakeholders).

should construct thoughtful, responsible legal oversight mechanisms for all predictive health analytics.

IV. RECOMMENDATIONS

At its core, this Article is a call to action. The policymaking and scientific communities must not ignore the potential risks of predictive health analytics. Just as the growth of genetic testing elicited robust academic and policy debates, so too should the burgeoning phenomenon of predictive health analytics. Effective legal and policy interventions are needed to safeguard the rights of individuals. This part recommends changes to the HIPAA Privacy and Security Rules and to the ADA. It also advocates for the implementation of other oversight and self-regulation mechanisms.

A. *Legal Interventions*

Legislators should modify the laws that establish privacy and antidiscrimination mandates. This section focuses on the federal HIPAA Privacy and Security Rules and the ADA, though states could make similar changes to parallel state statutes.²⁵⁰

1. The HIPAA Privacy and Security Rules

As I have argued in other works,²⁵¹ Congress and the Department of Health and Human Services should expand HIPAA's definition of "covered entity."²⁵² The need for change has become more urgent in light of the growing use of predictive health analytics. The federal law and regulations could use the language of a much broader Texas privacy statute as a model:

"Covered entity" means any person who:

(A) for commercial, financial, or professional gain, monetary fees, or dues, or on a cooperative, nonprofit, or pro bono basis, engages, in whole or in part, and with real or constructive knowledge, in the practice of assembling, collecting, analyzing, using, evaluating, storing, or transmitting protected health information. The term includes a business associate, health care payer, governmental unit, information or computer

250. See Sharona Hoffman, *Medical Privacy and Security*, in *THE OXFORD HANDBOOK OF U.S. HEALTHCARE LAW* 267, 274–78 (I. Glenn Cohen et al. eds., 2017) (discussing state privacy laws); *Disability Discrimination: State Law*, *WORKPLACE FAIRNESS*, <https://www.workplacefairness.org/disability-discrimination-state-law> [<https://perma.cc/L5K4-Y7W7>] (listing each state's disability discrimination laws).

251. See HOFFMAN, *supra* note 29, at 74 (advocating for a broader definition of "covered entity").

252. Hoffman & Podgurski, *supra* note 165, at 360–63.

management entity, school, health researcher, health care facility, clinic, health care provider, or person who maintains an Internet site;

(B) comes into possession of protected health information;

(C) obtains or stores protected health information under this chapter; or

(D) is an employee, agent, or contractor of a person described by Paragraph (A), (B), or (C) insofar as the employee, agent, or contractor creates, receives, obtains, maintains, uses, or transmits protected health information.²⁵³

Adoption of such language would require a parallel modification to the definition of “health information.”²⁵⁴ “Health information” should be expanded to mean:

any information, recorded in any form or medium, that relates to the past, present, or future physical or mental health or condition of an individual, including health predictions, the provision of healthcare to an individual, or the past, present, or future payment for the provision of healthcare to an individual.²⁵⁵

Expanding the definitions of “covered entities” and “health information” would not prevent healthcare providers from contracting with business associates, such as LexisNexis or Acxiom, to conduct predictive health analytics so long as they did so for purposes of treatment, payment, or healthcare operations.²⁵⁶ It also would not prevent data brokers from accessing much of the data they use, such as Facebook posts, shopper loyalty program records, or voter registration data.²⁵⁷

Nevertheless, the change would provide patients with several important benefits. First, it would prevent the newly covered entities from disclosing health predictions to third parties without the data subject’s consent. Because

253. TEX. HEALTH & SAFETY CODE ANN. § 181.001(b)(2) (Westlaw through 2019 Regular Session of 86th Legis.).

254. “Health information” currently means:

any information, whether oral or recorded in any form or medium, that –

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

(B) relates to the past, present, or future physical or mental health or condition of an individual, the provision of health care to an individual, or the past, present, or future payment for the provision of health care to an individual.

42 U.S.C. § 1320d(4) (2012); 45 C.F.R. § 160.103 (2018).

255. HOFFMAN, *supra* note 29, at 75.

256. *See supra* note 173 and accompanying text.

257. *See supra* Sections I.C–D.

these predictions would constitute health information about individuals' future physical or mental health conditions, data brokers and other commercial enterprises could not sell them for financial gain to marketers, employers, and other interested parties without permission. Second, upon request, the newly covered entities would be bound to inform data subjects of all disclosures made concerning their protected health information.²⁵⁸ Third, the newly covered entities would have to comply with the security mandates of the HIPAA Security Rule.²⁵⁹ They therefore would be prohibited from storing health information and health predictions about individuals using sloppy or minimal security measures that do not adequately deter hacking.

In addition, the HIPAA Privacy and Security Rules should include a private cause of action.²⁶⁰ Because of budgetary constraints, government enforcement is often anemic.²⁶¹ Furthermore, the rules do not provide aggrieved parties with monetary relief if they have suffered an injury resulting from a privacy breach.²⁶² The proposed HIPAA changes could meaningfully enhance data subjects' privacy protections and rights.

2. The Americans with Disabilities Act

I have also previously argued for a broadening of the ADA's definition of "disability," and I renew my call for this change here.²⁶³ The ADA's "regarded as" provision protects only individuals who are "being regarded as [currently] having . . . an impairment" from discrimination.²⁶⁴ Congress should revise the "regarded as" provision of the ADA to include individuals who "are perceived as likely to develop physical or mental impairments in the future."²⁶⁵

This change would prohibit employers and other parties from discriminating against individuals because of disease predictions.²⁶⁶ It follows logically from GINA, which forbids discrimination based on a specific type of predictive data—that is, genetic information.²⁶⁷ In the era of predictive health

258. 45 C.F.R. §§ 164.520(b)(1)(iv), .528 (2018).

259. *See id.* §§ 164.302–.318; *supra* notes 165–66 and accompanying text.

260. *See* HOFFMAN, *supra* note 29, at 78–79; *supra* notes 170–72 and accompanying text (discussing the absence of a private cause of action).

261. Morgan Leigh Tendam, *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, 421–22 (2018).

262. *See* HOFFMAN, *supra* note 29, at 75–76.

263. *See* Sharona Hoffman, *Big Data and the Americans with Disabilities Act*, 68 HASTINGS L.J. 777, 787–88 (2017) [hereinafter Hoffman, *Big Data*].

264. 42 U.S.C. § 12102(1)(C) (2012).

265. Hoffman, *Big Data*, *supra* note 263, at 787.

266. Hoffman, *New Discrimination*, *supra* note 205, at 94–96.

267. *See supra* Section II.C.1.

analytics, there is no justification for retaining a discrepancy between GINA and the ADA. Predictive models can forecast a myriad of health problems.²⁶⁸ These include inherited diseases such as heart conditions, some forms of Alzheimer's disease, and many more.²⁶⁹ GINA would not cover such predictions because they are not based on genetic tests or family histories.²⁷⁰ The law should not protect people only when the prediction of a future disease is rooted directly in genetic information. With respect to antidiscrimination mandates, genetic exceptionalism²⁷¹ no longer makes sense.

B. *Other Oversight Mechanisms*

Academics and other experts have begun building a literature about the legal and ethical implications of predictive health analytics only in recent years.²⁷² It is a long way from reaching the proportions of the genetic testing literature. Moreover, existing legal literature has paid little attention to long-term predictive health analytics, which raises significant concerns about psychological harms and discrimination.²⁷³ Legal and bioethics scholars should no more ignore these risks than they did the similar risks of genetic testing.²⁷⁴ What follows is a brief discussion of potential oversight improvements for the predictive health analytics industry and medical professionals.

1. Guidelines and Validation

A few papers have undertaken the development of initial guidelines for predictive health analytics.²⁷⁵ For example, a panel of seventeen experts proposed the following guiding principles in 2016:

1. Data Barriers: Establish mechanisms within the scientific community to support data sharing for predictive model development and testing.
2. Transparency: Set standards around e-HPA validation based on principles of scientific transparency and reproducibility.

268. See *supra* Sections I.B–C.

269. See *supra* note 94–96 and accompanying text.

270. See *supra* Section II.C.1.

271. Genetic exceptionalism is the belief that genetic information is special and should be treated differently from other health data. See James P. Evans & Wylie Burke, *Genetic Exceptionalism. Too Much of a Good Thing?*, 10 GENETICS MED. 500, 500–01 (2008).

272. See Cohen & Graver, *supra* note 185, at 446.

273. See *supra* Sections III.A–B.

274. See *supra* Section II.B.

275. See, e.g., Ruben Amarasingham et al., *Consensus Statement on Electronic Health Predictive Analytics: A Guiding Framework To Address Challenges*, 4 ELECTRONIC GENERATING EVIDENCE & METHODS 1, 1–2 (2016); Cohen et al., *supra* note 6, at 1139–40; Parikh et al., *supra* note 240, 811–12 (proposing five criteria “for evaluation and regulation of predictive algorithms”).

3. Ethics: Develop both individual-centered and society-centered risk-benefit approaches to evaluate e-HPA.
4. Regulation and Certification: Construct a self-regulation and certification framework within e-HPA.
5. Education and Training: Make significant changes to medical, nursing, and paraprofessional curricula by including training for understanding, evaluating, and utilizing predictive models.²⁷⁶

The scholars that have pondered predictive health analytics all agree that transparency and oversight are of critical importance.²⁷⁷ They recommend the establishment of industry-wide validation and certification mechanisms implemented by the Joint Commission, certifiers overseen by the FDA, independent institutional review boards, or other third parties.²⁷⁸

Experts have developed a variety of techniques to assess learning algorithms and predictive models.²⁷⁹ A popular method for estimating prediction error is cross-validation.²⁸⁰ Another method to assess statistical accuracy is the bootstrap method.²⁸¹ However, these techniques constitute internal validation that reuses the data with which the learning algorithm was trained.²⁸² It is even more important for researchers to engage in external validation of learning algorithms in the field, using real patients under the same conditions as those intended for the algorithm's post-approval use.²⁸³ Such validation, preferably at multiple sites and institutions, should ensure that the

276. Amarasingham et al., *supra* note 275, at 2, 9. "e-HPA" is electronic health predictive analytics.

277. *See id.* at 5–9; Cohen et al., *supra* note 6, at 1142–43; Marks, *supra* note 13 (manuscript at 31–32).

278. *See* Amarasingham et al., *supra* note 275, at 2, 5–9 (recommending the creation of a self-regulation framework); Cohen et al., *supra* note 6, at 1142–43 (recommending certification by outside third parties like the Joint Commission); Cortez, *supra* note 30, at 84–85 (discussing FDA-run approval systems); Marks, *supra* note 13 (manuscript at 30) (discussing independent IRB oversight).

279. *See* TREVOR HASTIE, ROBERT TIBSHIRANI & JEROME FRIEDMAN, *THE ELEMENTS OF STATISTICAL LEARNING: DATA MINING, INFERENCE, AND PREDICTION* 219–57 (2d ed. 2009) (discussing model assessment and selection).

280. *Id.* at 241–49 ("Probably the simplest and most widely used method for estimating prediction error is cross-validation.").

281. *See id.* at 249–54.

282. *See* Douglas G. Altman et al., *Prognosis and Prognostic Research: Validating a Prognostic Model*, 338 *BRITISH MED. J.* 432, 432–35 (2009); Noorazrul Yahya et al., *Independent External Validation of Predictive Models for Urinary Dysfunction Following External Beam Radiotherapy of the Prostate: Issues in Model Development and Reporting*, 120 *RADIOTHERAPY & ONCOLOGY* 339, 339–40 (2016).

283. *See* Altman et al., *supra* note 282, at 432; Yahya et al., *supra* note 282, at 339; Raeesa Gupte, *AI in Cancer Detection: Are We There Yet?*, *CLINICAL LAB MANAGER* (May 8, 2019), <https://www.clinicallabmanager.com/technology/ai-in-cancer-detection--are-we-there-yet--252> [<https://perma.cc/HW2W-G4U5>] (emphasizing the importance of randomized clinical trials that compare AI outcomes with physician diagnoses in the absence of AI).

algorithm's predictive capability generalizes to the true target population.²⁸⁴ Oversight bodies should consist of predictive health analytics and validation experts who can be trusted to scrutinize proposed assessment methods and ensure that they are appropriate.²⁸⁵

The recommendations offered thus far are sound, and experts should continue to develop and augment them in order to furnish policymakers with proposals that are as detailed and evidence-based as possible. Oversight and quality control could prevent many erroneous predictions and save clinicians and patients considerable angst.

2. Self-Regulation

Healthcare professionals should adopt their own safeguards in order to minimize the hazards of long-term disease predictions for patients, as they did in the case of genetic testing. To that end, physicians should receive training concerning long-term predictive health analytics so that they understand the extent to which it can be limited and uncertain. They should also counsel and educate patients before disclosing troubling health predictions to them. A process akin to genetic counseling would be very useful.²⁸⁶ Patients should understand the advantages and disadvantages of learning about their disease risks and be able to make informed decisions; they should not rush into obtaining long-term health predictions without carefully thinking through the potential for psychological harm, discrimination, and other adverse consequences.²⁸⁷

In addition, professional organizations should develop practice guidelines regarding when it is appropriate to employ predictive health analytics and the extent to which clinicians should rely upon it.²⁸⁸ They should thus follow the precedent set in the arena of genetic testing.²⁸⁹ For example, practice guidelines might recommend that clinicians refrain from obtaining certain types of predictions about children.²⁹⁰ They might also suggest which interventions should and should not be implemented in response to predictions of suicidal ideation, clinical depression, opioid addiction, or other ailments.

284. See *supra* Section III.C (discussing erroneous outcomes).

285. See Ewout W. Steyerberg & Frank E. Harrell Jr., *Prediction Models Need Appropriate Internal, Internal-External, and External Validation*, 69 J. CLINICAL EPIDEMIOLOGY 245, 245–47 (2016) (discussing the importance of external validation).

286. See *supra* notes 181–84 and accompanying text (discussing genetic counseling).

287. See *supra* notes 181–84 and accompanying text.

288. See *supra* notes 176–80 and accompanying text (discussing practice guidelines for genetic testing).

289. See *supra* notes 176–80 and accompanying text.

290. See *supra* notes 131–34 and accompanying text (discussing genetic testing of children).

Some industries have recognized the need for self-regulation and principled approaches to predictive analytics. For example, in March 2019, Google announced that it would launch an Advanced Technology External Advisory Council to provide ethics oversight and outside input regarding its development of artificial intelligence.²⁹¹ However, just days later, Google scrapped the Council because of protests regarding some of its members.²⁹²

Hopefully, other ethics initiatives will be more enduring and successful.²⁹³ Education, counseling, practice guidelines, and expert review could go far in maximizing the benefits and minimizing the risks of predictive health analytics.

CONCLUSION

We should not be blinded by enthusiasm for long-term predictive health analytics or be naively seduced by technologies with impressive names like “artificial intelligence” and “machine learning.” There is certainly much to be gained from prudent use of new predictive capabilities. However, the technologies come with significant risks of psychological harm, privacy violations, and discrimination, among others. Moreover, predictive models and learning algorithms are often flawed and produce erroneous outcomes. Many of these potential harms were previously considered and addressed in the context of genetic testing. Rather than leave a regulatory void, scientists and policymakers should adopt similar approaches for long-term predictive health analytics. This Article has proposed just a few legal and nonlegal interventions designed to enhance data subjects’ privacy rights, antidiscrimination protections, and ability to make informed decisions about obtaining disease predictions. However, many more minds must tackle the challenges of predictive health analytics and develop mechanisms to enhance the integrity and benefits of this technology. Ignoring the potential perils and unintended

291. Jillian D’Onfro, *Google Launches Advisory Council To Help Company Question Assumptions on Ethical AI*, FORBES (Mar. 26, 2019), <https://www.forbes.com/sites/jilliandonfro/2019/03/26/google-launches-advisory-council-to-help-company-question-assumptions-on-ethical-ai/#db487d779420> [https://perma.cc/TDK9-E2AF].

292. Jillian D’Onfro, *Google Scraps Its AI Ethics Board Less than Two Weeks After Launch in the Wake of Employee Protest*, FORBES (Apr. 4, 2019), <https://www.forbes.com/sites/jilliandonfro/2019/04/04/google-cancels-its-ai-ethics-board-less-than-two-weeks-after-launch-in-the-wake-of-employee-protest/#6b58a8d06e28> [https://perma.cc/HCSZ-QCHN]. The controversial members were Kay Coles James, President of the Heritage Foundation, a conservative think-tank, and Dyan Gibbens, CEO of a drone company called Trumbull Unmanned. *See id.*

293. *See, e.g., Axon AI and Policing Technology Ethics Board*, AXON, <https://www.axon.com/info/ai-ethics> [https://perma.cc/P4WM-P4MB]. Axon produces technology and weapon products for law enforcement, including the taser. *See* Stephen Nellis, *Taser Changes Name to Axon in Shift to Software Services*, REUTERS (Apr. 5, 2017), <https://www.reuters.com/article/us-usa-taser/taser-changes-name-to-axon-in-shift-to-software-services-idUSKBN177265> [https://perma.cc/K2TD-J9RF].

2019]

GENETIC TESTING

163

consequences of long-term predictive analytics is imprudent and could cost society dearly.

