Michigan Telecommunications and Technology Law Review

Volume 23 | Issue 1

2016

Privacy and Accountability in Black-Box Medicine

Roger Allan Ford
University of New Hampshire School of Law

W. Nicholson Price II
University of Michigan Law School, wnp@umich.edu

PRIVACY AND ACCOUNTABILITY IN BLACK-BOX MEDICINE

Roger Allan Ford† and W. Nicholson Price II‡

Cite as: Roger Allan Ford & W. Nicholson Price II,
Privacy and Accountability in Black-Box Medicine,
23 MICH. TELECOM. & TECH. L. REV. 1 (2016).
This manuscript may be accessed online at repository.law.umich.edu.

Black-box medicine—the use of big data and sophisticated machine-learning techniques for health-care applications—could be the future of personalized medicine. Black-box medicine promises to make it easier to diagnose rare diseases and conditions, identify the most promising treatments, and allocate scarce resources among different patients. But to succeed, it must overcome two separate, but related, problems: patient privacy and algorithmic accountability. Privacy is a problem because researchers need access to huge amounts of patient health information to generate useful medical predictions. And accountability is a problem because black-box algorithms must be verified by outsiders to ensure they are accurate and unbiased, but this means giving outsiders access to this health information.

This article examines the tension between the twin goals of privacy and accountability and develops a framework for balancing that tension. It proposes three pillars for an effective system of privacy-preserving accountability: substantive limitations on the collection, use, and disclosure of patient information; independent gatekeepers regulating information sharing between those developing and verifying black-box algorithms; and information-security requirements to prevent uninten-

† Associate Professor of Law, University of New Hampshire School of Law; Faculty Fellow, Franklin Pierce Center for Intellectual Property.

‡ Assistant Professor of Law, University of Michigan Law School; Affiliated Faculty, Petrie-Flom Center for Health Law Policy, Biotechnology, and Bioethics, Harvard Law School.

For helpful comments and conversations, we are indebted to Andrew Selbst, Anna Slomovic, Bob Gellman, Christo Wilson, Christopher Millard, Deborah Hurley, Dissent Doe, Felix Wu, Frank Pasquale, Hank Greely, Jacob Sherkow, Janine Hiller, Jay Kesan, Jennifer Berk, Margot Kaminski, Mark Lemley, Maya Bernstein, Melissa Goldstein, and Rebecca Eisenberg, and to participants at the Yale Information Society Project's Conference on Unlocking the Black Box, the Stanford Center for Law and the Biosciences Workshop, the thirteenth Works in Progress in Intellectual Property (WIPIP) Colloquium at the University of Washington, and the ninth Privacy Law Scholars Conference at George Washington University. Cassandra Simmons provided excellent research assistance.

tional disclosures of patient information. The article examines and draws on a similar debate in the field of clinical trials, where disclosing information from past trials can lead to new treatments but also threatens patient privacy.

INTRODUCTION

Medicine is an unpredictable science. A treatment that provides a miraculous recovery for one patient may do nothing for the next. A new chemotherapy drug may extend patient lives by two years on average, but that average consists of some patients who live many years longer and some patients whose lives are not extended at all, or even are shortened. And with new drugs costing more and more money, personalizing medicine is increasingly important, so that doctors can predict disease risk and choose treatments tailored to individual patients.

Medicine's unpredictability has a simple cause. The human body is one of the most complex systems in existence, with endless genetic variations, biological pathways, protein expression patterns, metabolite concentrations, and exercise patterns (to name just a few of the dozens of variables) affecting each person differently. And only a few of these variables are well-understood by scientists. When a drug doesn't work or a patient develops a rare disease, the reason could be some genetic variation or metabolite concentration or environmental difference—or several of these variables acting together in ways doctors will likely never understand.

Black-box medicine—the use of big data and sophisticated machine-learning techniques in opaque medical applications—could be the answer.[1]

1. *See, e.g.,* W. Nicholson Price II, *Black-Box Medicine,* 28 HARV. J.L. & TECH. 419 (2015) (defining black-box medicine); Ruben Amarasingham et al., *Implementing Electronic*

A scientist alone is unlikely to discover the precise combination of variables that makes a drug work or not. But with enough data, a machine-learning algorithm would have no trouble finding a predictive correlation. Using datasets of genetic and health information, then, researchers can uncover previously unknown connections between patient characteristics, symptoms, and medical conditions. And these connections promise to yield new diagnostic tests and treatments and to enable individually tailored medical decisions.

Big-data techniques are only as powerful as the input data and the methods used to analyze that data. Health care is especially ripe for a big-data revolution, though, because of the sheer quantity of data available: researchers can obtain an endless variety of data points from literally millions of patients. And because assembling and analyzing such large-scale datasets is becoming easier and cheaper by the hour, many different researchers, from both industry and the academy, are using data for everything from guiding choices between different drugs to best allocating scarce hospital resources among different patients.[2]

The sheer scale and scope of health data available to researchers, and the sensitivity of that data, lead to two related but opposing problems. The first problem is algorithmic accountability. Biological systems are so complex, and big-data techniques are so opaque, that it can be difficult or impossible to know if an algorithmic conclusion is incomplete, inaccurate, or biased. And these problems can arise due to data limitations, analytical limitations, or even intentional interference. Researchers or government agencies can sometimes validate an algorithm's conclusions, but doing so can be expensive and difficult, and can require access to the same extensive medical data from which the conclusions were drawn.

The second problem is privacy. Medical information can be some of the most private information that exists, and black-box medicine requires access to a *lot* of that information. It also creates new information, like predictions based on the models developed with big data. And this information may be used in ways that harms individuals, whether through marketing, sales to others, or discrimination in employment, insurance, or other decisions. Even

---

Health Care Predictive Analytics: Considerations And Challenges, 33 HEALTH AFF. 1148 (2014) (describing big-data predictions in health); Xiaoqian Jiang et al., *Calibrating Predictive Model Estimates to Support Personalized Medicine*, 19 J. AM. MED. INFORMATICS ASSOC. 263 (2012) (discussing predictive analytics and personalized medicine); Joseph A. Cruz & David S. Wishart, *Applications of Machine Learning in Cancer Prediction and Prognosis*, 2 CANCER INFORMATICS 59 (2007) (describing machine-learning approaches to cancer prediction). Janine S. Hiller, *Healthy Predictions? Questions for Data Analytics in Health Care,* 53 AM. BUS. L.J. 251 (2016).

2. *See* VIKTOR MAYER-SCHÖNBERGER & KENNETH CUKIER, BIG DATA: A REVOLUTION THAT WILL TRANSFORM HOW WE LIVE, WORK, AND THINK (2013) (describing many examples of how large-scale datasets are making differences in day-to-day life).

when it is not used in these ways, its collection, disclosure, and use can infringe individual autonomy and decisional privacy.

These two problems presented by black-box medicine are interrelated because efforts to reduce one will usually make the other worse. The solution to the accountability problem is to validate black-box models, but that requires access to more information, which can exacerbate the privacy problem. And the solution to the privacy problem is to limit the amount of information researchers, companies, and the government can use and to which they have access, but that can make it harder to validate models and easier to hide or overlook algorithmic problems. Algorithms need to be validated to ensure high-quality medicine, but at the same time, a data free-for-all would eviscerate patient privacy.

Solutions to the accountability and privacy problems, then, must consider the broader effects on black-box medicine. There are three pillars to an effective verification system that respects patient privacy. The first pillar is a system of limitations on the collection, use, and dissemination of medical data, so that data gathered and used to develop and verify black-box algorithms is not also used for illegitimate purposes. The second pillar is a system of independent gatekeepers to govern access to, and transmission of, patient data, so that government and independent researchers can verify big-data models. And the third pillar is robust information-security provisions, so that unintended outsiders cannot obtain, use, or disseminate patient data. The design of these verification systems can draw on the ongoing debate over the disclosure of clinical-trial data, which has addressed related issues of how to promote data sharing without sacrificing patient privacy. These verification systems are critical if black-box medicine is to live up to its promise without sacrificing patient privacy.

This article is structured in four parts. Part I provides background, describing the promise and genesis of black-box medicine. Parts II and III describe the two fundamental problems, with Part II discussing the challenge of algorithmic accountability and Part III discussing the challenge of protecting patient privacy. Part IV provides a structure to reconcile privacy and accountability.

I. Black-Box Medicine

Health care teems with complex problems. How should we treat a new cancer?[3] Which patients are most likely to benefit from organ transplants?[4]

3. *See* Janet E. Dancey et al., *The Genetic Basis for Cancer Treatment Decisions*, 148 Cell 409 (2012).

4. *See* Neil Mehta et al., *Identification of Liver Transplant Candidates with Hepatocellular Carcinoma and a Very Low Dropout Risk: Implications for the Current Organ Allocation Policy*, 19 Liver Transplantation 1343 (2013); Mark J. Russo et al., *Local Allocation of Lung Donors Results in Transplanting Lungs in Lower Priority Transplant Recipients*, 95 Ann. Thoracic Surgery 1231 (2013).

How should hospital systems use scarce resources to best promote patient health without wasting money on ineffective treatments?[5] Some answers can be gleaned from basic scientific research, clinical trials, and real-world experience. But many more answers lie beyond the scope of these processes, because they are too slow to provide answers, too expensive, or insufficiently nuanced.

Black-box medicine provides a shortcut, letting us discover and use answers to complex problems without fully understanding those problems or the answers themselves. This Part examines that shortcut. First, it describes the promise of black-box medicine. Then it discusses two technological trends—the growing volume of available health-care information and the growing power of methods to analyze that information—that have made black-box medicine possible.

A. *The Promise of Black-Box Medicine*

Black-box medicine can help solve complex medical problems by bringing to bear the power of big data. By using machine-learning algorithms to analyze massive amounts of individual medical data—medical big data—researchers can discover connections between specific patient attributes and specific symptoms, diseases, or treatments. The promise of black-box medicine, then, is that medical decisions can become personalized, predicting disease and tailoring diagnostics and treatment to individual patients.

Black-box techniques can answer several distinct kinds of medical questions. Some of these questions concern how to best allocate scarce health-care resources. For instance, we know some factors that affect which patients are most likely to benefit from organ transplants, and those factors, among others, are considered in maintaining organ-transplant priority lists.[6] But there are many other factors that influence the likelihood of success, and at least some of those factors are hidden in the reams of data about transplant-patient outcomes. Finding these patterns could help better allocate scarce organs to the patients most likely to benefit from them. Similarly, patterns hidden in existing health-care data could identify patients who need urgent care or who are most likely to develop some future disease, allowing physicians to intervene quickly and avoid greater harm in the future.[7]

Other questions concern how best to treat a particular patient. Many medical treatments affect different patients differently; the same treatment, given to two patients suffering from the same disease, may cure one patient

5. *See* Ruben Amarasingham et al., *Allocating Scarce Resources in Real-Time to Reduce Heart Failure Readmissions: A Prospective, Controlled Study*, 22 BRIT. MED. J. QUALITY & SAFETY 998 (2013).

6. Russo et al., *supra* note 4; Mehta et al., *supra* note 4.

7. *See, e.g.*, Amarasingham et al., *supra* note 5; DANIEL GARTNER, OPTIMIZING HOSPITAL-WIDE PATIENT SCHEDULING: EARLY CLASSIFICATION OF DIAGNOSIS-RELATED GROUPS THROUGH MACHINE LEARNING (2015).

but have no effect for the other.[8] A black-box algorithm could guide treatment decisions by predicting that one drug might work better than another for a specific patient, or might have fewer side effects, or that the patient would likely respond best to a particular dose on a particular schedule.[9] Such algorithms could eliminate the need for physicians to experiment with different drugs, saving significant time and money.

A different set of questions concerns how to quickly and efficiently diagnose diseases and conditions. Many diseases are simple to diagnose. Blood tests, for instance, can conclusively diagnose many viral illnesses; hypertension can be easily diagnosed with a sphygmomanometer. But others are more difficult. Some diseases and conditions simply don't have reliable tests; others so closely resemble each other, or include so many subtypes, that it is hard to tell different diseases or conditions apart.[10] And sometimes a patient may have underlying risk factors that may develop into a problematic condition.[11] Black-box algorithms could help diagnose these uncertain diseases and conditions. A black-box model might, for instance, identify the specific genes that predict who will develop a disease or condition; or it might tell physicians, earlier than they could otherwise tell, which of two similar diseases a patient has.[12]

Some of these benefits can be obtained through other means; others are unique to black-box medicine. But even when other means are available, black-box algorithms could significantly reduce health-care costs by eliminating the need to perform other, costly tests or to waste time on ineffective

---

8. *See* Margaret A. Hamburg & Francis S. Collins, *The Path to Personalized Medicine*, 363 NEW ENG. J. MED. 301 (2010) (describing personalized medicine).

9. *See* U.S. Food & Drug Admin., *Personalized Medicine*, http://www.fda.gov/scienceresearch/specialtopics/personalizedmedicine/default.htm (describing the goal of providing "the right patient with the right drug at the right dose at the right time.").

10. *See, e.g.*, Konstantina Kourou et al., *Machine learning applications in cancer prognosis and prediction*, 13 COMPUTATIONAL & STRUCTURAL BIOTECHNOLOGY J. 8 (2015) (reviewing the application of machine-learning techniques to the problem of classifying different types of cancer); Steven I. Sherman et al., *Augmenting pre-operative risk of recurrence stratification in differentiated thyroid carcinoma using machine learning and high dimensional transcriptional data from thyroid FNA*, 33 J. CLINICAL ONCOLOGY 6044 (2015) (reporting a study using machine learning to classify thyroid cancer tumors); Vivek Subbiah & Razelle Kurzrock, *Universal Genomic Testing Needed to Win the War Against Cancer: Genomics IS the Diagnosis,* 2 JAMA ONCOLOGY 719 (2016).

11. *See* Yiran Guo et al., *Machine learning derived risk prediction of anorexia nervosa*, 9 BMC MED. GENOMICS, no. 4, 2016, at 1 (reporting on an effort to use machine-learning techniques to assess patients' risk of developing the eating disorder anorexia nervosa based on genetic data).

12. *See e.g.*, Graziella Orrù et al., *Using Support Vector Machine to Identify Imaging Biomarkers of Neurological and Psychiatric Disease: A Critical Review*, 36 NEUROSCIENCE & BIOBEHAVIORAL REV. 1140 (2012); Elaheh Moradi et al., *Machine Learning Framework for Early MRI-Based Alzheimer's Conversion Prediction in MCI Subjects*, 104 NEUROIMAGE 398 (2015); Zhi Wei et al., *Large Sample Size, Wide Variant Spectrum, and Advanced Machine-Learning Technique Boost Risk Prediction for Inflammatory Bowel Disease*, 92 AM. J. HUM. GENETICS 1008 (2013).

treatments. And they could lead to better health outcomes as inappropriate treatments or dangerous side effects are avoided.

To be sure, there is no guarantee that black-box medicine will live up to its promise; significant challenges will have to be overcome if black-box algorithms are to become a routine and accepted part of health care for most patients. Our goal in this article is not to argue that black-box medicine will be a cure-all, or even a solution to rising health-care costs. Instead, we observe that the trend of researchers making use of patient health information, including in black-box algorithms, is only likely to accelerate. And that acceleration will create several challenges for policy makers, including the challenge of protecting patient privacy while promoting algorithmic verification.

B. *The Genesis of Black-Box Medicine*

Black-box medicine is becoming possible because of two related technological trends. One is that the health-care system generates ever-larger amounts of health data about individual patients. The other is that new analytic tools like machine learning make it possible to analyze those vast troves of health data to find underlying patterns. These tools don't reveal the causes of those patterns; they simply reveal predictions or recommendations on which doctors can rely. A doctor might know, then, that a particular treatment is the best option for a patient with certain genetic markers without knowing how or why those markers matter. This is the black box of black-box medicine: decisions can be based on opaque algorithmic analysis of dozens or hundreds of variables, with no theories to explain the results.[13]

Big data in health. The first trend is the increasing amounts of health data available to researchers. Black-box medicine would not be possible without this mass of health data, which is rapidly growing both because routinely collected data is more broadly available and because new forms of measurement create data that didn't previously exist.[14]

Data that have long been collected are newly accessible for several reasons. Routine clinical visits are now recorded in electronic health records

13. Price, *supra* note 1, at 433–34. Not all machine-learning algorithms are fully black-box; some are truly opaque, others yield patterns so complex as to be largely uninterpretable, and some methods may in fact be interpreted. *Id.* And some may move from being opaque to being interpreted as reverse-engineering techniques advance. We focus here on the first two categories, both of which may be described as "black box." Greater privacy protections may be needed when using interpretable algorithms, since an output from an algorithm may reveal inputs and further compromise privacy. *See infra* note 101 (discussing a study reverse engineering an algorithm for predicting optimal Warfarin dosages).

14. *See* David W. Bates et al., *Big Data in Health Care: Using Analytics to Identify and Manage High-Risk and High-Cost Patients,* 33 HEALTH AFF. 1123 (2014).

rather than being relegated to paper files in doctors' offices.[15] Pharmacy records are now consolidated in electronic databases of prescription and fulfillment information.[16] Insurance claims are amassed in administrative claims databases that record episodes of care.[17] When all of these records were kept in paper, any effort to use them to develop or verify black-box algorithms would be prohibitively expensive or logistically overwhelming. Now that they are electronic, though, new forms of analysis become possible.[18]

New technologies are also generating new types of health data. Genetic sequencing may be the most important of these new technologies, because it reveals a vast number of inherited differences between individuals.[19] While genetic sequencing was, until recently, prohibitively expensive, it has rapidly dropped in price, reaching $1,245 per patient in October 2015.[20] As a result, hundreds of thousands of patients have had their entire genomes sequenced;[21] many more have been tested for specific genetic variations[22] or

15. *See* Julia Adler-Milstein et al., *Electronic Health Record Adoption In US Hospitals: Progress Continues, But Challenges Persist*, 34 HEALTH AFF. 2174, 2176 (2016) (finding that 75.2% of hospitals had adopted at least basic electronic health-records systems by 2014).

16. *See* Joy M. Grossman et al., *Transmitting and Processing Electronic Prescriptions: Experiences of Physician Practices and Pharmacies*, 19 J. AM. MED. INFO. ASS'N 353, 353 (2012) (An "important e-prescribing feature is the two-way electronic exchange of prescription data between physicians and pharmacies. Physicians can transmit new prescriptions directly from their e-prescribing systems into pharmacy information systems as well as respond to pharmacies' electronic renewal authorization.").

17. *See*, *e.g.*, Colin R. Cooke & Theodore J. Iwashyna, Using existing data to address important clinical questions in critical care, 21 CRIT. CARE MED. 886, 889 (2013) (describing administrative claims data).

18. *See* Peter B. Jensen et al., *Mining Electronic Health Records: Towards Better Research Applications and Clinical Care,* 13 NATURE REVIEWS GENETICS 395, 395 (2012) ("Databases in modern health centres automatically capture structured data relating to all aspects of care, including diagnosis, medication, laboratory test results and radiological imaging data. This transformation holds great promise for the individual patient as richer information, coupled with clinical decision support (CDS) systems, becomes readily available at the bedside to support informed decision making and to improve patient safety.").

19. *See* Wylie Burke & Bruce M. Psaty, *Personalized Medicine in the Era of Genomics*, 298 J. AM. MED. ASSOC. 1682 (2007) ("Enthusiastic predictions about personalized medicine have surrounded the sequencing of the human genome . . . . [G]enomics-based knowledge and tools promise the ability to approach each patient as the biological individual he or she is, thereby radically changing our paradigms and improving efficacy.").

20. *See The Cost of Sequencing a Human Genome*, NATIONAL HUMAN GENOME RESEARCH INSTITUTE, https://www.genome.gov/27565109/the-cost-of-sequencing-a-human-genome/ (last updated Jan. 15, 2016).

21. *See* Antonio Regalado, *EmTech: Illumina Says 228,000 Human Genomes Will Be Sequenced in 2014*, MIT TECHNOLOGY REVIEW (Sept. 24, 2014), https://www.technologyreview.com/s/531091/emtech-illumina-says-228000-human-genomes-will-be-sequenced-this-year/.

22. *See* Matthew B. Yurgelun et al., *Population-Wide Screening for Germline BRCA1 and BRCA2 Mutations: Too Much of a Good Thing?*, 33 J. CLIN. ONCOLOGY 3092 (2015) (describing genetic tests indicating predispositions toward breast or ovarian cancer).

have parts of their genomes sequenced.[23] Beyond genetic sequencing, other new technologies also generate new forms of health data, including measurements of gene expression[24] and screens for the presence and levels of various metabolites in the body.[25] And some of these tools are increasingly available directly to consumers, like 23andMe's gene-testing services.[26] New health data also comes from more prosaic sources. Personal activity trackers like those sold by Fitbit and Apple, for instance, record individuals' activity, providing new data that could be used in black-box medicine. Similarly, shopping patterns can predict health outcomes.[27]

When all of these categories of health data of different types are combined, from different sources, covering tens or hundreds of millions of patients, the result is enough data to reveal even subtle trends in health information. To be sure, there are substantial technological, economic, and legal hurdles in bringing together data from different sources into coherent records while ensuring data quality.[28] And even a carefully assembled dataset can imperfectly represent the patient population, since it may draw from a homogeneous or underinclusive population, or can reflect existing bias by physicians and other providers in diagnoses, prescriptions, tests or-

23. *See* Kiera Peikoff, *Fearing Punishment for Bad Genes*, N.Y. TIMES (Apr. 17, 2014), http://www.nytimes.com/2014/04/08/science/fearing-punishment-for-bad-genes.html (reporting, in 2014, that 700,000 Americans had their DNA sequenced, in whole or in part).

24. *See, e.g.*, Laura J. van 't Veer et al., *Gene Expression Profiling Predicts Clinical Outcome of Breast Cancer*, 415 NATURE 530, 530 (2002) ("Here we used DNA microarray analysis on primary breast tumours of 117 young patients, and applied supervised classification to identify a gene expression signature strongly predictive of a short interval to distant metastases in patients without tumour cells in local lymph nodes at diagnosis."); Christina D. Lyngholm et al., *Validation of a Gene Expression Profile Predictive of the Risk of Radiation-Induced Fibrosis in Women Treated with Breast Conserving Therapy*, 54 ACTA ONCOLOGICA 1665, 1665 (2015) ("[W]e examined the frequency and degree of late morbidity related to [breast cancer treatment] and demonstrated moderate to severe fibrosis (grade ii–iii) in the residual breast in 23% of the patients 7–20 years after treatment.").

25. *See* Omran Abu Aboud & Robert H. Weiss, *New Opportunities from the Cancer Metabolome*, 59 CLINICAL CHEMISTRY 138 (2013) (describing the use of metabolite screens in cancer treatment).

26. Using 450,000 customers' genetic information, 23andMe was able to pinpoint 15 genetic locations associated with clinical depression. *See* Craig L. Hyde et al., *Identification of 15 genetic loci associated with risk of major depression in individuals of European descent,* 48 NATURE GENET. 1031 (2016); Antonio Regalado, *23andMe Pulls Off Massive Crowdsourced Depression Study,* MIT TECH. REV., https://www.technologyreview.com/s/602052/23andme-pulls-off-massive-crowdsourced-depression-study/ (Aug. 1, 2016).

27. *See* Rebecca Robins, *Insurers Want to Nudge You to Better Health. So They're Data Mining Your Shopping Lists*, STAT NEWS (Dec. 15, 2015), http://www.statnews.com/2015/12/15/insurance-big-data/ ("Shopping at home-improvement stores, for instance, turns out to be a great predictor of mental health. If you suddenly stop shopping at Lowe's, your insurance company may suspect that you're depressed.").

28. *See* Price, *supra* note 13 (describing the challenges of gathering, cleaning, and linking data for black-box medicine).

dered, or notes taken.[29] But these hurdles are at least partially addressable and larger and more comprehensive health datasets are becoming available, if slowly, for analysis and innovation. For instance, as part of President Obama's Precision Medicine Initiative, the NIH is assembling a comprehensive dataset, including health records, questionnaire data, and genetic information, on a million volunteers; that uniform and comprehensive dataset will be a useful tool for developing better black-box algorithms.[30] Similarly, the Million Veteran Program aims to collect blood samples and health records from one million veterans for use in health research.[31]

New analytical tools. The second trend is the development of new tools for analyzing large datasets to find patterns. Raw data are not particularly useful for making health-care decisions; instead, the patterns and relationships that are embedded in data can provide information that drives decisions. So the critical question is how to find these underlying patterns, and the answer lies in learning algorithms being developed in the field of machine learning and artificial intelligence.[32] These algorithms are already being used in numerous contexts, medical and otherwise. For instance, Google's Image Search and Facebook's DeepFace facial-recognition algorithm both use black-box image-recognition algorithms to determine the content of images and identify faces.[33] Similar algorithms can find clusters of factors that predict the risk of developing a disease or benefiting from a specific drug.[34] These machine-learning algorithms come in many forms, including one technology, deep-learning neural networks, that is literally modeled on the human brain.[35]

---

29. *See, e.g.*, Sarah E. Malanga et al., *Big Data Neglects Populations Most in Need of Medical and Public Health Research and Interventions* (draft on file with authors); Jonas Lerman, *Big Data and Its Exclusions*, 66 STAN. L. REV. ONLINE 55 (2013); Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CAL. L. REV. 671 (2016); Pauline T. Kim, *Data-Driven Discrimination at Work,* 58 WM. & MARY L. REV. (forthcoming 2017); *see also infra* note 102 and accompanying text.

30. *See Precision Medicine Initiative Cohort Program – FAQ*, NAT'L INSTITUTES OF HEALTH (July 6, 2016), https://www.nih.gov/precision-medicine-initiative-cohort-program/precision-medicine-initiative-cohort-program-frequently-asked-questions.

31. *See Million Veteran Program*, U.S. DEPT. OF VETERANS AFF., http://www.research.va.gov/mvp/ (last visited Sept. 9, 2016).

32. For an overview of machine-learning techniques, *see* PETER FLACH, MACHINE LEARNING: THE ART AND SCIENCE OF ALGORITHMS THAT MAKE SENSE OF DATA (2012).

33. *See* Yaniv Taigman et. al, *DeepFace: Closing the Gap to Human-Level Performance in Face Verification*, RESEARCH AT FACEBOOK (June 24, 2014), https://research.facebook.com/publications/deepface-closing-the-gap-to-human-level-performance-in-face-verification/.

34. *See President's Council of Advisors on Science and Technology: Priorities for Personalized Medicine*, WHITEHOUSE.GOV 13, 55 (September 2008), https://www.whitehouse.gov/files/documents/ostp/PCAST/pcast_report_v2.pdf (providing examples of applications of algorithms in the personalized medicine context).

35. *Neural Network*, OXFORD ENGLISH DICTIONARY (online) (last visited Nov. 6, 2016).

While the details of different machine-learning algorithms are not important for this article, one common difficulty is: machine-learning algorithms are typically opaque, meaning the patterns found and used are not transparent to the user. Facebook's DeepFace, for instance, uses a neural network trained on four million images to analyze faces; while the network can tell whether two faces are the same with 97.35% accuracy, it cannot tell us *why* two faces are the same, or state with any intelligibility what features are used in that classification.[36] Such an opaque algorithm can provide answers, but can't explain those answers or provide any generalizable theories.

Opacity comes in different forms; an algorithm can be deliberately or inherently opaque. With a deliberately opaque algorithm, the developer chooses to keep secret specific details of the algorithm, such as the algorithm itself or the way it was developed. These deliberately opaque algorithms are increasingly common in everything from stock-market trading to credit ratings to Internet search results.[37]

Machine-learning tools like those used in black-box medicine are different. In many cases they generate algorithms that are unavoidably opaque.[38] These algorithms typically cannot identify the reasons for the patterns they find, due to the iterative processes by which the algorithms are developed. Neural networks, like those used in DeepFace, are modeled on the human brain, with artificial "synapses" that are followed or not, depending on the success or failure of a test. The eventual result of training those processes cannot be stated explicitly.[39] And even when patterns discovered by an algorithm can be stated, those patterns are typically far too complex to be of much use in understanding underlying mechanisms. Knowing that a combi-

---

36. Taigman, *supra* note 33, at 1, 4 (describing dataset, noting accuracy, and describing the six-layer neural network and connections between the layers).

37. FRANK PASQUALE, THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION 4 (2015) ("Credit raters, search engines, major banks, and the TSA take in data about us and convert it into scores, rankings, risk calculations, and watch lists with vitally important consequences. But the proprietary algorithms by which they do so are immune from scrutiny, except on the rare occasions when a whistleblower litigates or leaks.")

38. To be sure, black-box medicine can also be deliberately opaque, as when developers keep secret the way they develop algorithms, the data on which those algorithms are based, or the ways those algorithms were validated. This secrecy presents potential challenges for ensuring quality, as well as for the incentives and ability to develop new algorithms in the first place. *See* W. Nicholson Price II, *Patents, Big Data, and the Future of Medicine*, 37 CARDOZO L. REV. 1401 (2016) (discussing the incentive benefits and challenges of secrecy in the development of black-box medicine) [hereinafter Price, *Big Data*].

39. *See* Tom Simonite, *Facebook Creates Software that Matches Faces Almost as Well as You Do*, MIT TECH. REV. (Mar. 17, 2014) https://www.technologyreview.com/s/525586/facebook-creates-software-that-matches-faces-almost-as-well-as-you-do/ (reporting that Facebook's software "uses networks of simulated neurons to learn to recognize patterns in large amounts of data"); Taigman, *supra* note 33 (reporting that Facebook's algorithm involves more than 120 million parameters).

nation of 5,000 specific gene alleles predicts drug response in a lung tumor, for instance, tells us little about why that is.[40]

Although black-box techniques offer substantial potential to advance health care, the accompanying opacity presents a different kind of challenge from those faced by most medical technologies. It is much harder to be confident that something works and is safe when we cannot understand *how* it works or *why* it is safe. Physicians, patients, insurers, and regulators will be skeptical, and some of this skepticism will be justified. Not all black-box techniques will work, and not all that do work will work well. But some will, and the potential benefits from those that do work are large. Maximizing these benefits requires confronting the accountability challenge: how can we ensure that algorithms adopted in practice are high-quality and reliable?

II. The Accountability Challenge

The first problem that arises from the growth of black-box medicine is algorithmic accountability. Patients, health-care providers, and insurers must be able to trust in the quality of black-box medicine before they will rely on it. And this trust is fundamentally difficult to establish due to the inherent opacity of black-box algorithms. Since no one knows—or can know—exactly how black-box algorithms work, traditional means of showing medical quality largely fail. Users can't rely on scientific understanding of an algorithm, as they frequently do with conventional treatments and diagnostic methods, and clinical trials fit poorly with complex, opaque algorithms except at the very broadest level. Black-box medicine, then, demands new methods of verification, like computational verification performed by independent third parties.

A. *The Need for Verification*

Verification is the process of proving that a medical device, treatment, or diagnostic test works and performs its intended function.[41] Black-box algorithms need verification, just like other health-care goods, for three major reasons. First, health-care goods are classic credence goods: consumers can't independently evaluate their quality and so must accept on faith that a treat-

40. *See* Hojin Moon et al., *Ensemble Methods for Classification of Patients for Personalized Medicine with High-Dimensional Data,* 41 Artif. Intell. Med. 197, 203–04 (2007).

41. For diagnostic tests, verification is often divided into three parts: analytical validity (whether a test accurately measures what it purports to measure), clinical validity (whether what the test measures accurately reflects an underlying clinical characteristic), and clinical utility (whether the test can be used to usefully guide care). *See Levels of Evidence for Cancer Genetics Studies*, National Cancer Institute, http://www.cancer.gov/publications/pdq/levels-evidence/genetics (last updated July 25, 2012). The third is most important for black-box medicine, since the first two will often be unknown.

ment or test works as intended.[42] When someone is sick, takes a drug, and gets better, maybe the drug worked—or maybe the patient would have gotten better just as quickly, or more quickly, without the drug. And since it is usually difficult or impossible to determine in individual cases which possibility is most likely, the field of medicine uses other methods, like randomized double-blind clinical trials, to determine efficacy.[43]

Second, verification is essential for black-box medicine (and other health-care goods) because health-care markets are poorly suited to weeding out low-quality products, even when reliable information is available. Health-care markets are unusually complex, with no single, knowledgeable consumer having an incentive to select high-quality products. Instead, doctors typically select treatments, insurers pay for those treatments, and patients benefit from them.[44] Patients want to get well and to minimize out-of-pocket costs and inconvenience; doctors want to provide beneficial interventions but face complex incentives regarding the volume and cost of care;[45] and insurers want to decrease their own costs, while avoiding costlier later illnesses—unless patients are likely to have switched insurers by then. Doctors may also be susceptible to automation bias, trusting algorithms even when they haven't been proven reliable. These overlapping and conflicting incentives, and the complexity of the existing mechanisms for selecting and paying for care, mean that the health-care market cannot, on its own, easily choose high-quality care.[46]

---

42. *See, e.g.*, Uwe Dulleck & Rudolf Kerschbamer, *On Doctors, Mechanics, and Computer Specialists: The Economics of Credence Goods,* 44 J. ECON. LIT. 5, 5-6 (2006) ("Goods and services where an expert knows more about the quality a consumer needs than the consumer himself are called credence goods."); W. Nicholson Price II, *Regulating Complex and Black-Box Medical Algorithms* (draft manuscript on file with authors) ("For patients especially, medical algorithms, like many other medical technologies, are 'credence goods' whose efficacy must generally be taken on faith—or, more accurately, on the word of those with more knowledge.").

43. *See* Susan White Junod, FDA and Clinical Trials: A Short History, FDA, http://www.fda.gov/AboutFDA/WhatWeDo/History/Overviews/ucm304485.htm (last updated July 7, 2014). ("Although several kinds of randomized controlled trial methodologies can be useful to researchers and regulators, ultimately, it was the randomized, double-blinded, placebo controlled experiment which became the standard by which most other experimental methods were judged, and it has often subsequently been referred to as the 'gold' standard for clinical trial methodology.").

44. A rich literature addresses the varying incentives of different players in the health-care system. *See, e.g.*, HANDBOOK OF HEALTH ECONOMICS (Anthony Cuyler & Joseph Newhouse eds., 2000); INCENTIVES AND CHOICE IN HEALTH CARE (Frank Sloan & Hirschel Kasper eds., 1st ed. 2008).

45. In pure fee-for-service models, doctors face incentives to increase the volume and cost of care; in newer models, those incentives may be tempered by limits on payment or revenue-sharing schemes.

46. *Cf.* Saurabh Bhargava et al., *Do Individuals Make Sensible Health Insurance Decisions? Evidence from a Menu with Dominated Options* (National Bureau of Economic Research, Working Paper No. 21160, 2015), *available at* http://www.nber.org/ (finding that a

Third, verification is essential for black-box medicine because of its inherent opacity, which makes algorithms harder to evaluate and more likely to be of highly variable quality. Some black-box algorithms and products will work well, some will work poorly, and some will not work at all, depending on whether they rely on real underlying relationships that reflect actual biology or spurious correlations that happen to arise in a particular database or slice of a large database. And this variability is exacerbated by barriers to entry that should fall significantly over time. At this time, developing black-box algorithms requires access to large amounts of patient information that must be collected, at great expense, along with substantial programming expertise and computational infrastructure.[47] But eventually, these costs will fall as data is combined into large datasets and computational tools become standardized. Indeed, developing lower-cost tools is much of the point of black-box medicine. But with low entry barriers comes the likelihood of low-quality products developed by low-quality developers.

These low-quality algorithms could be quite bad for patients. Some potential errors would be relatively benign, as when an algorithm incorrectly suggests diagnostic testing based on an elevated risk that was never actually elevated. Unnecessary diagnostic tests are not cost-free, but are less likely to cause serious harms.[48] But other potential errors would be more harmful, as when an algorithm recommends the wrong drug or the wrong dose of the right drug.[49] To be sure, such mistakes already happen in great numbers.[50] But the promise of black-box medicine is to reduce them, not to perpetuate or increase them. Black-box medicine could also reflect embedded bias or discriminatory rules, whether introduced through biased datasets or through

---

majority of employees studied chose dominated health insurance plan options, resulting in substantial excess spending).

47. *See* Price, *supra* note 1, at 449 (describing the costs of developing black-box medicine).

48. *See, e.g.*, Joann G. Elmore et al., *Ten-Year Risk of False Positive Screening Mammograms and Clinical Breast Examinations*, 338 N. ENGL. J. MED. 1089 (1998) (finding that over a 10-year period, each woman undergoing regular mammograms and clinical breast exams faced a cumulative 49.1% risk of a false positive diagnosis); David W. Dowdy et al., *Is Scale-Up Worth It? Challenges in Economic Analysis of Diagnostic Tests for Tuberculosis*, 8 PLOS MED., 1–3 (2011) (estimating the false-positive costs of a scaled-up tuberculosis diagnostic test); Timothy J. Wilt & Philipp Dahm, *PSA Screening for Prostate Cancer: Why Saying No Is a High-Value Health Care Choice*, 13 J. NAT'L COMPREHENSIVE CANCER NETWORK 1566 (2015) (arguing against prostate-specific-antigen screening for prostate cancer because it results in little benefit and substantial cost).

49. *See* Kit Huckvale et al., *Smartphone Apps for Calculating Insulin Dose: A Systematic Assessment*, 13 BMC MED. 106 (2015) (noting pervasive errors in smartphone apps used to calculate the dosage of insulin for diabetics). Similarly, if an algorithm suggested the wrong drug for a developing malignant tumor, the time lost to ineffective treatment could make the cancer harder to treat eventually.

50. The literature on medical error is voluminous; perhaps the highest-profile summary is INSTITUTE OF MEDICINE, TO ERR IS HUMAN: BUILDING A SAFER HEALTH SYSTEM (Linda T. Kohn, Janet M. Corrigan, & Molla S. Donaldson eds., 2000).

biases in algorithmic development.[51] A robust verification system could help avoid these errors and biases, which would otherwise be difficult for consumers to avoid.

B. *Verification by Clinical Trials*

Though clinical trials are the most common means of verifying medical devices, treatments, and diagnostic tests, they face substantial challenges in the context of black-box medicine. Normally, the FDA uses clinical trials to validate new medical treatments. And clinical trials could be usefully applied to some types of black-box medicine. For example, in a situation in which a black-box algorithm recommends which of two cancer drugs a patient should take, a group of patients could be randomly separated; some would have their treatment regime assigned according to the algorithm, and others according to the standard of care. Improvements in results for the algorithm-assigned patients would point to an algorithm that improves on the standard of care.

There are limits, though, to how useful clinical trials can be in verifying black-box algorithms. Clinical trials are expensive, slow, and blunt tools for their ordinary role of ensuring new drugs are safe and effective; they're even worse when it comes to verifying fast-moving black-box algorithms. Clinical trials for new drugs cost hundreds of millions, or billions, of dollars.[52] Although clinical trials to support approval of new medical devices—including diagnostic tests—are far cheaper than those to support approval of a new drug, they still cost $1 to $10 million or more.[53] Clinical trials to demonstrate a new use for an already-approved drug—a likely, and desira-

---

51. *See supra* note 28 and accompanying text; *see also* Kate Crawford, *Artificial Intelligence's White Guy Problem*, N.Y. TIMES (June 25, 2016), http://www.nytimes.com/2016/06/26/opinion/sunday/artificial-intelligences-white-guy-problem.html (describing biases embedded in artificial-intelligence algorithms).

52. Clinical trials for new drugs take an average of five years and are usually estimated to cost an average of several hundred million to almost three billion dollars, though the exact figures are hotly disputed. *See, e.g.*, Joseph A. DiMasi et. al, *The Price of Innovation: New Estimates of Drug Development Cost*, 22 J. HEALTH ECON., 151 (2003) (estimating that the cost of obtaining FDA preapproval is $802 million over 5 years); Christopher P. Adams & Van V. Brantner, *Estimating The Cost Of New Drug Development: Is It Really $802 Million?*, 25 HEALTH AFF., 420 (2006) (estimating the cost of obtaining FDA preapproval to vary between $500 million and $2 billion over 52 months); Yeveniy Feyman, *Opinion: Shocking Secrets of FDA Clinical Trials Revealed*, FORBES (Jan. 24, 2014), http://www.forbes.com/sites/theapothecary/2014/01/24/shocking-secrets-of-fda-clinical-trials-revealed/ (updating DiMasi's $802 million estimate in 2003 to $1.3 billion in 2009); *but see* Aylin Sertkaya et. al, *Examination of Clinical Trial Costs and Barriers for Drug Development*, U.S. DEPT. HEALTH & HUMAN SERVICES, Table 1 (July 25, 2014), https://aspe.hhs.gov/report/examination-clinical-trial-costs-and-barriers-drug-development (estimating that the cost of Phase I, II, and III clinical trials are approximately $20 to $70 million).

53. For an overview of the medical device approval pathway, *see*, Aaron V. Kaplan et al., *Medical Device Development From Prototype to Regulatory Approval*, 109 CIRCULATION 3068 (2004).

ble, outcome of black-box medicine—are substantially more expensive.[54] These cost hurdles would sharply limit the potential of black-box medicine to generate cheaper medical innovation based on already-existing data.

The slowness of clinical trials would also make it hard to harness the ability of black-box medicine to quickly find new patterns in ever-expanding datasets. Clinical trials typically take three to five years.[55] Machine-learning algorithms, though, can find useful relationships in weeks, days, or hours, and can update their predictions in real-time as new data is added. Relying on multi-year clinical trials to verify black-box algorithms would substantially delay their usefulness; relying on a new clinical trial each time an algorithm was updated would quickly prove unworkable.

Even if clinical trials were cheaper and quicker, they are poorly matched to the problem of verifying black-box interventions that are truly precise or personalized. A clinical trial works by assembling a cohort of comparable patients, treating a randomized subset of that cohort with the intervention to be tested, and observing whether that subset improves measurably. This works for, say, a drug that is supposed to work for everyone with lung cancer, because one need only assemble a cohort of lung-cancer patients. And it works, though with more effort, for drugs with more complex indications, such as a drug designed to help diabetic children with lung cancer, since those cohorts are harder, though not impossible, to assemble. But black-box medicine can identify interventions that are tailored to single individuals, not groups with similar indications. And without the ability to randomize across a set of similar patients and observe different outcomes, there is no way to conduct a clinical trial of an algorithm that predicts individual responses of individual patients. One possibility is a meta-trial, in which members of a cohort subset are treated according to the predictions of a black-box algorithm (whatever the treatment for each individual) and members of a different subset are untreated, or treated according to a standard protocol, but this is far from the typical trial designed to evaluate a particular intervention.

To be sure, randomized double-blind trials are second to none at demonstrating that an intervention is safe and effective, and some physicians will undoubtedly look skeptically at black-box algorithms that haven't been subjected to clinical trials. Other tools for verifying an algorithm, like the computational-verification tools discussed next, are a close approximation. But they will never be able to demonstrate causation in the same way that a randomized double-blind clinical trial can. And clinical trials could still be useful in some circumstances to verify black-box predictions or interventions. For instance, black-box predictions of which prostate cancers are rapidly progressing and which are indolent could be used to guide care of a random selection of men being screened. This would verify the strength of

54. *See supra* note 52.
55. *See supra* note 52.

the prediction at low cost.[56] But the practical obstacles to subjecting each algorithm to a clinical trial would be enormous, and the likely benefits of doing so would, in many cases, be small. Regulations should be cautious, then, before requiring a clinical trial for each black-box algorithm.

Despite these difficulties, the FDA appears to be turning toward the clinical-trial model to verify black-box algorithms. In October 2014, the FDA issued a Draft Guidance for Industry on the regulation of laboratory-developed tests, which are developed and used in individual laboratories instead of packaged and sold more broadly.[57] Under the FDA's broad definition of medical devices, it would regulate many forms of black-box medicine as laboratory-developed tests.[58] And the Draft Guidance explained that it considered to be higher risks "those devices that claim to enhance the use of a specific therapeutic product, through selection of therapy, patient population, or dose, but which are not included in the therapeutic product labeling (e.g., devices developed by laboratories that claim to predict who will respond to a therapy approved for use in a larger population)."[59] Such targeting diagnostics are the heart of black-box medicine. Their classification as higher-risk means that they would need to be validated by clinical trials before use, implicating the problems described above.

The FDA's approach is not yet set in stone and may not end up applying to all black-box medicine. And clinical trials could help demonstrate quality in some black-box algorithms, especially as they go from development into medical practice. They could also help persuade skeptical physicians, patients, and insurers that black-box algorithms are trustworthy. But requiring clinical trials for every black-box algorithm would be a substantial obstacle for algorithm developers. If the FDA maintains the broad, relatively inflexible approach suggested in the Draft Guidance, it risks significantly slowing the development of high-quality black-box medicine.

---

56. *See* Shazia Irshad et al., *A Molecular Signature Predictive of Indolent Prostate Cancer*, 5 SCI. TRANSLATIONAL MED. 202ra122 (2013) ("Gleason score prostate tumors can be distinguished as indolent and aggressive subgroups on the basis of their expression of genes associated with aging and senescence. Using gene set enrichment analysis, we identified a 19-gene signature enriched in indolent prostate tumors.").

57. FDA, DRAFT GUIDANCE FOR INDUSTRY, FOOD AND DRUG ADMINISTRATION STAFF, AND CLINICAL LABORATORIES: FRAMEWORK FOR REGULATORY OVERSIGHT OF LABORATORY DEVELOPED TESTS (LDTs), (October 3, 2014), *available at* http://www.fda.gov/downloads/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm416685.pdf [hereinafter LDT Draft Guidance].

58. A full discussion of FDA regulation of black-box medicine is complex, involving FDA's limitation on regulating the practice of medicine, the difference between laboratory-developed tests and more broadly sold *in vitro* diagnostic kits, and the practical realities of FDA's informal power. Such a discussion is outside the scope of this work.

59. LDT Draft Guidance, *supra* note 57, at 27.

C. *Computational Verification*

Computational verification is the principal alternative to clinical trials for verifying the quality of black-box algorithms. While clinical trials are poorly suited to validating black-box algorithms, computational verification based on patient data can demonstrate the quality of black-box algorithms while preserving their power, flexibility, and speed. It does this by harnessing the same big-data techniques used to develop black-box algorithms to demonstrate the validity of those algorithms.

Computational verification requires duplicating the critical results of a black-box algorithm using different input data, different analytical methods, or both.[60] If one team develops a black-box model predicting, say, which patients are most likely to benefit from a chemotherapy drug, then maybe those predictions reflect genuine patterns found in nature—or maybe they're artifacts of the particular data or methods used to generate the model. But if a second team, using different data or methods, develops a model that comes to similar conclusions, then those conclusions are more likely to reflect genuine patterns on which doctors, patients, and insurers can rely. Just how similar is enough is a hard question. It is impossible to compare directly two opaque black-box models, since one can't, for instance, ask whether 120 million variables are weighted the same between the models. But if representative patient information is fed to both models and gives similar predictions, then we can be confident that the models are based on natural phenomena or, at least, the same underlying flaw.

Performing computation verification is conceptually simple, but implementing it is a complex undertaking for both technical and legal reasons. On the technical side, choosing and obtaining access to the right data and method presents several challenges that will take researchers years to iron out. And on the legal side, regulators overseeing black-box medicine face several important questions about how to verify black-box algorithms, including when verification should be encouraged or required, who should perform that verification, and what data and analytical techniques should be used for verification. A full discussion of these questions is well beyond the scope of this article, but we offer some preliminary observations on how regulators should think about these questions.[61]

First, though new drugs must be approved before they can be marketed, it is less clear that pre-market approval should be required for black-box medicine. Pre-market approval would help ensure that black-box treatments

60. Third parties (or the FDA) could similarly evaluate the procedural quality of an algorithm's development, including the relevant expertise of its developers, the quality of training data, and the like. This procedural validation would require companies to disclose their developmental parameters, however, which raises intellectual-property and regulatory questions beyond the scope of this paper. *See* Price, *supra* note 42 (discussing companies' incentives to keep development parameters as trade secrets).

61. For a more detailed consideration of these issues, *see* Price, *supra* note 42.

and diagnostics are safe and effective, but would also impose significant costs that may retard the development of black-box algorithms or even prevent them from getting off the ground in the first place. And unlike new pharmaceutical treatments, the safety concerns of black-box medicine are more attenuated. While pharmaceutical treatments subject patients to chemicals that may or may not work and may or may not have dangerous side effects, black-box algorithms largely predict risks of developing disease or help physicians decide how best to treat a disease using existing, FDA-approved drugs.[62] So pre-market approval might be too restrictive a requirement.

Other approaches could obtain many of the same benefits as pre-market approval without the downsides. For instance, the FDA could require companies to obtain verification within a specific period of time of introducing a commercial product based on a black-box algorithm, or upon achieving a specific sales volume. Or it, or the marketplace, could provide incentives to verify algorithms without imposing specific requirements. An independent certification program, for instance, might guide decisions by physicians, insurers, and the Centers for Medicare and Medicaid Services when considering whether to prescribe and pay for products based on black-box models. Such a certification program could be run by independent nonprofits or international organizations, or even the FDA itself.

Second, verification should, in most cases, be done by independent third parties rather than by the original developer of a black-box model. Original developers are poorly suited to perform validation: though they have the necessary expertise and relevant data, they are also likely to repeat any errors in the initial development. They also face substantial conflicts of interest, since they are precisely the parties who benefit from verification. Developers of black-box algorithms have their own incentive to verify their own models internally, since doing so improves their quality, which will be reflected in better products. But the point of verification, as seen by patients, physicians, and insurers, is to avoid the errors and conflicts of interest that can only be overcome by independent third parties. Independent verification could be performed by the FDA or NIH, but those agencies generally lack the necessary expertise in big-data management and computer programming to develop their own black-box algorithms; nor is doing so their expected

62. It is possible for a black-box model to create safety problems. Consider, again, a hypothetical model predicting which patients are most likely to benefit from a chemotherapy drug. A truly terrible black-box model might make predictions that are the opposite of correct, directing doctors to give the drug to those who would be harmed and to avoid giving the drug to those who would benefit. Such a model would make things worse off than they were before, but only marginally; without the model, doctors might prescribe the drug indiscriminately both to those who would benefit and those who would be harmed. And, of course, if the model didn't work, it might be quickly rejected by physicians and insurers.

strength or mission.[63] Instead, validation is likely best undertaken by third parties with enough expertise to develop algorithms on their own—other health-care companies, independent research organizations, or academic researchers.

Encouraging these third parties to perform independent verification is a key goal, which could be driven by a variety of incentive mechanisms.[64] In particular, mandatory revenue sharing may be appropriate when a third-party health-care company or research organization verifies a black-box algorithm; similarly, a third party that debunks a black-box algorithm could collect a bounty paid by the original developer or out of an FDA-administered fund. Grants and prizes to fund verification research could also provide useful incentives, especially for academic institutions capable of verifying black-box algorithms. Of course, original developers would have strong incentives to block third-party verification, when doing so could reveal valuable information or risk lucrative revenue streams. In addition to the incentives discussed above, then, the FDA would likely need to require companies to facilitate independent verification.[65]

Third, as explained above, independent verification should be based on different data or analytical methods than those used to develop the relevant black-box algorithms. Even with these differences in data and methods, however, third parties conducting independent verification would still need broad access to data used to develop black-box algorithms, because the number of entities capable of developing (and validating) black-box algorithms is larger than the number capable of assembling large, high-quality health-care datasets. Sharing and pooling health information while maintaining the ability of independent verification to avoid false correlations and data artifacts can be accomplished in various ways, like sectioning large pools of health information into subsets used for development and verification.[66]

---

63. The FDA's mission statement recognizes roles in promoting public health and encouraging health-care innovation:

> FDA is responsible for protecting the public health by assuring the safety, efficacy and security of human and veterinary drugs, biological products, medical devices, our nation's food supply, cosmetics, and products that emit radiation.
>
> FDA is also responsible for advancing the public health by helping to speed innovations that make medicines more effective, safer, and more affordable and by helping the public get the accurate, science-based information they need to use medicines and foods to maintain and improve their health.

ABOUT FDA: WHAT WE DO, FDA.gov (last updated Dec. 7, 2015) http://www.fda.gov/AboutFDA/WhatWeDo/.

64. *See* Price, *Big Data*, *supra* note 38, at 137–50.

65. *See also infra* Part IV.B (discussing the role of independent gatekeepers to regulate information sharing for verification).

66. The question of fully independent data is complex. On the one hand, data from independent sources may be best at ensuring patterns reflect underlying biological truths instead of artifacts of the original dataset. On the other hand, maintaining independent datasets

The many complex details aside, this is the heart of the accountability challenge: ensuring the quality of black-box algorithms requires sharing the health information used to develop the algorithms in the first place.[67] But with that information sharing comes a substantial second challenge: the privacy of those whose data are gathered in the first place. The next Part details that challenge.

III. The Privacy Challenge

The second problem that arises from the growth of black-box medicine is protecting patient privacy. Health care presents substantial privacy challenges because health information is, unusually, both private and public: it is personal information about which individuals have strong privacy preferences, yet is used by others for numerous valuable applications. Black-box medicine elevates these challenges in several ways. It uses far more information than traditional health-care applications; it requires comprehensive access to, and wide distribution of, health information; and it generates new health information that can present its own privacy problems. The resulting losses of patient privacy, then, can cause significant harms, both from inappropriate uses of health information and from effects on patient autonomy and decisional privacy.

A. *Health Information and Patient Privacy*

Privacy challenges arise in numerous contexts, from employment history and consumer credit to family dynamics and sexual relationships. But the health-care context presents some of the hardest privacy challenges, due to the nature of health information.

Privacy challenges arise when information has a dual nature: when it is both sensitive or private enough for interested parties to demand privacy and

---

necessarily reduces the size of those datasets, increases the cost of developing and maintaining them, and increases the likelihood of uncompensated biases. *See* Price, *Big Data*, *supra* note 38, at 110–13. One solution, though the details are beyond the scope of this article, might be a single dataset as large and comprehensive as possible, which developers could independently segregate into their own "training" and "test" sets. *See* José Ramón Cano et al., *On the Combination of Evolutionary Algorithms and Stratified Strategies for Training Set Selection in Data Mining*, 6 Appl. Soft Comput. 323 (2006). Different developers would divide the data differently, mimicking the existence of fully independent datasets.

67. For some of the technical challenges involved in sharing and analyzing large-scale health data, *see* Guy Haskin Fernald et al., *Bioinformatics Challenges for Personalized Medicine*, 27 Bioinformatics 1741 (2011). For a discussion of the intellectual-property and incentive issues surrounding access to health-care data, *see, e.g.*, Barbara J. Evans, *Sustainable Access to Data for Postmarketing Medical Product Safety Surveillance under the Amended HIPAA Privacy Rule*, 24 Health Matrix 11 (2014) (discussing access to FDA's Sentinel database); Price, *Big Data*, *supra* note 38, at 131–35 (discussing the incentives for secrecy of proprietary health data).

yet valuable or useful enough for others to want or need access.[68] Information that is private but has no value to anyone else, like someone's inner monologue or get-rich-quick fantasies, doesn't present significant privacy challenges because such information can just remain private with no cost or harm. Information that is valuable but not private, like a company's SEC filings or the United States Code, can likewise be made available at no privacy cost. It's when information falls between these categories—when there is a mismatch between the supply and demand for information—that privacy problems arise.

Health information is a classic example of this phenomenon. Many categories of health information are considered private by most people, for a wide variety of reasons. Some, like the details of a patient's mental-health treatment or a diagnosis of a sexually transmitted infection, can affect a patient's personal or professional relationships. Some, like a patient's weight, or photos from a colonoscopy, are just embarrassing or inappropriate when distributed in the wrong contexts.[69] And some, like a patient's genetic profile, can be used to discriminate in insurance, employment, and other decisions that may turn on someone's propensity to develop a medical condition.

At the same time, this information also has value to others. Some of this value is obvious: doctors need it to make diagnoses and provide treatment, while insurers need it to process claims. Sometimes this value is more attenuated, but still socially desirable, like when someone seeks information about her partner's sexual health or when a public-health service tracks the spread of a food-borne illness. And sometimes information is put to uses that have private value but also social costs, like when an employer discriminates on the basis of a protected medical condition or (arguably) when a pharma-

68. Privacy scholars have debated in recent years whether health information, or any form of information, is inherently "sensitive," such that it necessarily presents privacy problems, or whether the privacy implications of a given category of information depend more on the context in which that information is collected, used, or disclosed, or the privacy harms that stem from that collection, use, or disclosure. *See, e.g.*, HELEN NISSENBAUM, PRIVACY IN CONTEXT: TECHNOLOGY, POLICY, AND THE INTEGRITY OF SOCIAL LIFE (2010); Paul Ohm, *Sensitive Information*, 88 S. CAL. L. REV. 1125 (2015); Kirsten Martin & Helen Nissenbaum, *Measuring Privacy: An Empirical Test Using Context to Expose Confounding Variables* (forthcoming). We do not take sides in this debate. Whether health information is inherently sensitive, or whether patients consider any particular category of information to be sensitive, it is clear that many types of health information present privacy issues, meaning that their collection, use, or disclosure can lead to privacy harms. *See infra* Part III.C. Some forms of health information undoubtedly present greater privacy issues than others—as the joke goes, everyone knows someone in therapy, while no one has ever met anyone who has visited a proctologist—but in the aggregate, health information presents enough privacy issues that protecting patient privacy is a critical issue for researchers developing black-box algorithms.

69. *See generally* DANIEL J. SOLOVE, UNDERSTANDING PRIVACY 158–61 (2008); Nissenbaum, *supra* note 68, at 221 ("[I]nformation . . . online, particularly if taken out of a local context, may be embarrassing or cause . . . harm.").

ceutical company markets a drug to people suffering from a medical condition.

The law has taken steps to protect patient privacy in some, but not all, circumstances. In the United States, health privacy is governed, in large part, by the Health Insurance Portability and Accountability Act (HIPAA) and the Department of Health and Human Services' implementing Privacy Rule.[70] Under the Privacy Rule, most health-care providers, insurance companies, and information clearinghouses may not use or disclose identifiable health information unless that information falls within one of several listed categories.[71] These include using information to provide care or obtain payment, using information for quality-improvement efforts, disclosing information in response to a legal requirement, and disclosing or using information with a patient's consent.[72] The Privacy Rule also provides that in most contexts, other than use to provide care, the amount of information disclosed must be the minimum necessary, preventing most bulk disclosures of information.[73] Deidentified data—data that has been stripped of its personally identifiable information—is not, however, governed by the Privacy Rule, leading to its frequent use in health research.[74] This has upsides and downsides: it makes more information available to researchers, but deidentified data from different sources is much harder to combine into unified databases.[75] Furthermore, deidentified data can often be reidentified, leading to new privacy losses.[76]

---

70. Health Insurance Portability and Accountability Act of 1996, Pub. L. No. 104-191, 110 Stat. 1936 (1996); 45 C.F.R. pts. 160, 164 (2016).

71. *See* 45 C.F.R. §§ 160.103, 164.502 (2016) (defining covered entities and protected health information, and prohibiting unauthorized use or disclosure of protected health information, respectively).

72. *See* 48 C.F.R. § 164.506 (2016).

73. *See* 48 C.F.R. § 164.502(b) (2016).

74. *See* 45 C.F.R. § 164.514 (2013) (exempting de-identified data).

75. *See* INST. OF MED., BEYOND THE HIPAA PRIVACY RULE: ENHANCING PRIVACY, IMPROVING HEALTH THROUGH RESEARCH 177-79 (Sharyl J. Nass, Laura A. Levit & Lawrence O. Gostin eds., 2009) ("[B]ecause datasets from multiple sources cannot be linked to generate a more complete record of a patient's health history without a unique identifier, such datasets often are of minimal value to researchers and are not frequently used.").

76. *See, e.g.*, Paul Ohm, *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization*, 57 UCLA L. REV. 1701, 1716 (2010) ("About fifteen years ago, researchers started to chip away at the robust anonymization assumption, the foundation upon which this state of affairs has been built. Recently, however, they have done more than chip away; they have essentially blown it up, casting serious doubt on the power of anonymization, proving its theoretical limits and establishing . . . the easy reidentification result."); Paul M. Schwartz & Daniel J. Solove, *The PII Problem: Privacy and a New Concept of Personally Identifiable Information*, 86 N.Y.U. L. REV. 1814, 1841–45 (2011) ("Technology increasingly enables the combination of various pieces of non-PII to produce PII. . . . The more information about a person that is known, the more likely it becomes that this information can be used to identify that person or to determine further data about her. When aggregated, information has a way of producing more information, such that de-identification of data becomes more difficult. Thus, it becomes possible to look for overlap in the data and then to link up different bodies of data."); Felix T. Wu, *Defining Privacy and Utility in Data Sets*, 84 U. COLO. L. REV. 1117,

Other laws also play roles in governing health privacy. In the federal system, for instance, the Genetic Information Nondiscrimination Act of 2008 prohibits discrimination in employment or health-insurance decisions on the basis of genetic information.[77] With employer-provided health plans, the Employee Retirement Income Security Act of 1974 (ERISA) also plays a role, preempting many state provisions that might otherwise affect patient privacy.[78] And state tort law has long imposed duties of confidentiality on physicians and others treating patients, as have some state statutes, either in specific domains or comprehensively.[79]

B. *The Privacy Challenge of Black-Box Medicine*

Health information presents difficult privacy challenges, but those challenges are multiplied by the growth of black-box medicine, for at least four reasons.

Mass quantities of health information. First, black-box medicine requires enormous quantities of health information, expanding greatly the

1127-28 (2013) (discussing well-publicized instances in which researchers showed how to re-identify individuals in supposedly anonymous data). For computer-science literature on identification of individuals using purportedly anonymous data, *see, e.g.*, Arvind Narayanan & Vitaly Shmatikov, *Robust De-anonymization of Large Sparse Datasets*, *in* PROCEEDINGS OF THE 2008 IEEE SYMPOSIUM ON SECURITY AND PRIVACY 111–125 (2008) ("[V]ery little auxiliary information is needed [to] deanonymize an average subscriber record from the Netflix Prize dataset."); Yves-Alexandre de Montjoye et al., *Unique in the Shopping Mall: On the Reidentifiability of Credit Card Metadata*, 347 SCI. 536 (2015) ("We study 3 months of credit card records for 1.1 million people and show that four spatiotemporal points are enough to uniquely reidentify 90% of individuals."); Latanya Sweeney, *k-Anonymity: A Model For Protecting Privacy*, 10 INT'L J. ON UNCERTAINTY, FUZZINESS & KNOWLEDGE-BASED SYS. 557 (2002) ("[I]n most of these cases, . . . remaining data can be used to re-identify individuals by linking or matching the data to other data or by looking at unique characteristics found in the released data."); Latanya Sweeney, *Simple Demographics Often Identify People Uniquely* (Carnegie Mellon Univ., Working Paper No. 3, 2000), *available at* http://dataprivacylab.org/projects/identifiability/paper1.pdf ("[C]ombinations of few characteristics often combine in populations to uniquely or nearly uniquely identify some individuals. Clearly, data released containing such information about these individuals should not be considered anonymous.").

77. Genetic Information Nondiscrimination Act of 2008, Pub. L. No. 110-233, 122 Stat. 881 (2008).

78. *See* Gobeille v. Liberty Mut. Ins. Co., 136 S. Ct. 936 (2016) (holding that ERISA preempts a Vermont law requiring health-care providers and insurers to report claims information for inclusion in a state-run database).

79. *See, e.g.*, CAL. CIV. CODE §§ 56-56.37 (West 2014) (imposing broad confidentiality duties on health-care providers and plans); N.Y. PUB. HEALTH LAW § 17 (LexisNexis 2016) (prohibiting nonconsensual disclosure of medical records of minors relating to abortion and STIs); Pennsylvania Drug and Alcohol Abuse Control Act, 71 PA. CONS. STAT. § 1690.108 (1972) (prohibiting nonconsensual disclosure of medical records related to treatment programs for alcohol or drug abuse); McCormick v. England, 494 S.E. 2d 431, 439 (S.C. Ct. App. 1997) (recognizing the tort of breach of confidentiality by a physician); Hammonds v. Aetna Cas. & Sur. Co., 243 F. Supp. 793, 803 (N.D. Ohio 1965) (awarding tort damages where insurer induced doctor to divulge confidential patient information, thus breaching the doctor's duty of loyalty owed to the patient).

amount of patient information that must be collected and used. This quantity problem arises in two separate dimensions: black-box medicine depends both on having access to medical information about many individuals—thousands or perhaps millions of patients—and on having access to many distinct data points about each individual. Having such massive amounts of data is key because black-box medicine works by finding subtle correlations between patient characteristics and medical diagnoses or treatments. But there are so many patient variables that could be relevant to any given medical condition—whether from genetic testing, bloodwork and other laboratory tests, environmental characteristics, or other sources of variation—that machine-learning algorithms must analyze enormous amounts of data to draw useful inferences. Otherwise, with so many variables, false correlations would arise simply by coincidence and overfitting.

Comprehensive health information. Second, black-box medicine can require access to comprehensive health information. Comprehensiveness comes in different forms. In one form, this means that a dataset cannot systematically exclude certain categories of patients, since doing so would introduce a significant source of error. In this sense, comprehensiveness is just a variation on the quantity problem described above. But more importantly, it means that data from many different providers must be shared and combined into larger, unified datasets. Since this requires transferring sensitive medical information from different providers to a central repository, it opens up avenues for malicious actors to gain access to information as it is being transmitted. And it amplifies the benefits of doing so, since a single attack can gain access to far more information than in a decentralized system.

Broad distribution of health information. Third, black-box medicine can require distribution of health information to numerous recipients, from doctors and other health-care providers to laboratories performing analysis and testing to researchers investigating potential correlations. Much of this distribution has not previously been required. In the traditional pre-big-data approach to health care, for instance, a doctor might learn from a published study about a new treatment for a condition. She could then try that treatment on her patients without conveying information about those patients to others. In the age of black-box medicine, though, the doctor may send a patient's genetic information to a third party to be analyzed using a black-box model. If that model indicates the optimal treatment, this process would give better health outcomes. But it would also mean that sensitive information is in more hands than under the traditional approach.

Both of these latter two reasons mean that in the age of black-box medicine, health information is far more often susceptible to interception or misuse. And this is true for a wide variety of actors, from criminals hacking into servers to marketers selling new treatments. With so much data being collected about so many people, and being transmitted to and possessed by

so many recipients, the risk is necessarily greater that data will be used for purposes beyond those originally intended.

Creation of new health information. Fourth, black-box medicine leads to the creation of new health information that wouldn't have existed otherwise, in the form of the precise inferences that it enables.[80] For instance, when researchers identify genes that are linked with specific diseases, like the BRCA1 and BRCA2 mutations that indicate a high risk of developing breast or ovarian cancer, then that finding creates a new kind of medical inference about people with that genetic profile. The BRCA mutations provide unusually strong correlations, but the point holds even in murkier circumstances: If a black-box model suggests that a patient is likely to develop heart disease or diabetes, or is unlikely to respond to a cheap medicine but likelier to respond to a costlier one, those probabilities and susceptibilities to treatment are relevant medical facts in which the patient, her physicians, and her insurers all have strong interests.

And much of this new information presents the same privacy issues as other kinds of health information. This is obviously true when the information has financial consequences, as when it can lead to discrimination in insurance or employment decisions. But even when money isn't at stake, information in the form of black-box predictions can lead to the same kinds of privacy harms as other forms of health information.[81]

Protecting privacy in black-box medicine, then, is a difficult challenge both because of the amount and nature of the information in play and because that information often must travel on distributed networks of healthcare providers, pharmaceutical companies, academic and government researchers, and others. And since these characteristics of black-box medicine are necessary to obtain its utility, solutions are unlikely to come simply from reducing the amount of information or limiting its distribution.

C. *Privacy Harms from Black-Box Medicine*

A tempting response to these difficulties is to conclude that privacy be damned, the benefits of black-box medicine are so promising that patients should get comfortable with the costs. There is some merit to this view. Concerns about health privacy may be attenuated in the context of black-box medicine, since the massive amounts of data needed mean that the odds that any one patient's information will ever be seen or used by any person are low. Still, there are reasons people consider health information to be especially sensitive. One reason, though not the only one, is that losses of medical privacy can lead to four kinds of harms.

---

80. *See* Roger Allan Ford, *Unilateral Invasions of Privacy*, 91 N.D. L. REV. 1075, 1088-90 (2016); *but see* Jeffery M. Skopek, *Privacy in Numbers?* (on file with authors) (arguing that predictions and inferences may not appropriately be considered as privacy violations).

81. *See infra* Part III.C.

One class of privacy harms arises when a privacy loss causes what Ryan Calo has called objective privacy harms—objective, real-world consequences from the collection, disclosure, or use of information.[82] Many of these harms are financial. If a patient's medical records indicate she is suffering from an expensive-to-treat illness, for instance, then an insurer might use that information to deny her coverage, or a potential employer might use it to deny her a job, or a scammer might use it to sell her a snake-oil cure. But privacy losses can also lead to non-financial objective privacy harms. Disclosure of medical records showing that someone suffers from a sexually transmitted infection, for instance, or that someone's child has an unexpected biological parent, is likely to hurt the subject's reputation or family relations, even if it leads to no immediate financial loss.[83] It is relatively uncontroversial that these privacy harms are genuine harms that reasonably could, in many cases, merit compensation; they are, however, probably the rarest category of privacy harms.

A second class of privacy harms consists of subjective privacy harms—harms that are perceived, internally, by an information subject but that have no immediate real-world consequences.[84] These can range from mild to severe and consist of numerous distinct kinds of suffering—discomfort, embarrassment, paranoia, mental pain. There is extensive psychological evidence that these kinds of feelings inflict genuine harm.[85] That is why intentional infliction of emotional distress is a cause of action, or why even the threat of unwanted physical contact is recognized as the tort of assault.[86] And losses of privacy in the health-care context are especially likely to cause this sort of privacy harm, since health information is so personal and sensitive.

---

82. M. Ryan Calo, *The Boundaries of Privacy Harm*, 86 IND. L.J. 1131, 1147–52 (2011).

83. *See, e.g.,* SOLOVE, *supra* note 69, at 174–79. For real-life examples of such privacy breaches, *see, e.g.*, Doe v. Medlantic Health Care Group, 814 A.2d 939, 947 (D.C. Ct. App. 2003) (finding that a hospital violated its duty of confidentiality by sharing Doe's HIV-positive status, which led to ostracism at work); Yath v. Fairview Clinics, 767 N.W.2d 34, 50 (Minn. Ct. App. 2009) (finding that a clinic employee violated his duty of confidentiality when he posted information about the plaintiff's STI status and extramarital affair on the Internet). Lawmakers have occasionally targeted disclosure of such information. *See, e.g.,* 42 U.S.C. §§ 290dd-3, 390ee-3 (imposing requirements on the disclosure of information relating to treatment programs for drug and alcohol abuse); Timothy S. Jost, *Constraints on Sharing Mental Health and Substance-Use Treatment Information Imposed by Federal and State Medical Records Privacy Laws in* INSTITUTE OF MEDICINE (US) COMMITTEE ON CROSSING THE QUALITY CHASM: ADAPTATION TO MENTAL HEALTH AND ADDICTIVE DISORDERS (Washington DC: National Academies Press, 2006) (describing state and federal restrictions on information sharing).

84. Calo, *supra* note 82, at 1142–47.

85. SOLOVE, *supra* note 69, at 174–79.

86. *Id.* at 176.

A third class of privacy harms arises when the loss of privacy deprives people of their dignity, personhood, or individual autonomy. Like those in the first two classes of privacy harms, these harms are personal rather than social, but they represent a group of harms that are more abstract than those in the first two categories. Privacy is valuable, in part, because it represents a respect for the capabilities of individuals to make decisions, develop relationships, experience emotions, and generally live autonomous lives. Without a private realm in which to exercise these and other fundamental human capabilities, there is a real risk that people will lose key elements of individual liberty.[87] Indeed, this is a central justification for the Supreme Court's cases on a constitutional right to decisional privacy.[88] Eliminating privacy in this personal realm, then, would threaten to deprive individuals of the ability to live autonomous lives and exercise fundamental human capabilities.

A fourth class of privacy harms, and the broadest category of harms, follows when the absence of privacy alters behavior in a way that hurts individuals or society. Privacy creates environments that foster cooperation, trust, and confidence; without privacy, people are likely to be far more guarded in their interactions, or to avoid interactions that would benefit society.[89] In *Jaffee v. Redmond,*[90] for instance, the Supreme Court recognized an evidentiary privilege for conversations between psychotherapists and patients in part because confidentiality can be necessary for therapy to work in the first place. The Court explained that "[e]ffective psychotherapy . . . depends upon an atmosphere of confidence and trust in which the patient is willing to make a frank and complete disclosure of facts, emotions, memories, and fears," so that "the mere possibility of disclosure may impede de-

87. This is an application of the capabilities approach, which posits that individual well-being is a function of individual ability to do and be those things that are valuable. For background on the capabilities approach, *see, e.g.*, MARTHA C. NUSSBAUM, CREATING CAPABILITIES: THE HUMAN DEVELOPMENT APPROACH (2011) (positing core capabilities of life, bodily health, bodily integrity, emotions, practical reason, control over one's environment, and so forth); MARTHA C. NUSSBAUM, WOMEN AND HUMAN DEVELOPMENT: THE CAPABILITIES APPROACH 72–80 (2000); AMARTYA SEN, COMMODITIES AND CAPABILITIES (1985); Alexander A. Boni-Saenz, *Personal Delegations,* 78 BROOK. L. REV. 1231, 1233–34 (2013). On human capabilities and privacy, *see* Martha C. Nussbaum, *Sex Equality, Liberty, and Privacy: A Comparative Approach to the Feminist Critique, in* INDIA'S LIVING CONSTITUTION: IDEAS, PRACTICES, CONTROVERSIES 242 (Zoya Hasan et al. eds., 2002).

88. Planned Parenthood of Southeastern Pa. v. Casey, 505 U.S. 833, 851 (1992) ("Our law affords constitutional protection to personal decisions relating to marriage, procreation, contraception, family relationships, child rearing, and education. . . . These matters, involving the most intimate and personal choices a person may make in a lifetime, choices central to personal dignity and autonomy, are central to the liberty protected by the Fourteenth Amendment.").

89. On this public-good nature of privacy, *see* Paul M. Schwartz, *Property, Privacy, and Personal Data*, 117 HARV. L. REV. 2055, 2084–90 (2004).

90. Jaffee v. Redmond, 518 U.S. 1, 15 (1996).

velopment of the confidential relationship necessary for successful treatment."[91]

Similar issues arise in myriad other contexts, even when the subject matter is not so personal. When two businesses sign a nondisclosure agreement, for instance, before negotiating a contract or entering into a joint project, the privacy environment created by the NDA allows the businesses to disclose confidential information—necessary for the negotiations or project to succeed—without sacrificing the value of the confidentiality. Without the promise of privacy, many of these valuable interactions—between doctor and patient or lovers or business partners—might never happen, or might be irredeemably tainted by exposure to the world. By recognizing and protecting privacy in these kinds of contexts, then, the law creates an environment that encourages and protects the formation of these partnerships in their most vulnerable times.[92]

IV. Reconciling Privacy and Accountability

The last two Parts explained the challenges that algorithmic accountability and patient privacy present for black-box medicine. These problems are not unsolvable. Similar problems in traditional medicine have been addressed, to greater or lesser success, by the FDA's drug-approval process and the HIPAA privacy rule. These problems are, however, harder to solve than they may seem at first glance, because efforts to combat one will usually make the other worse. Efforts to address accountability or privacy, then, must consider the effects on the other problem. This Part suggests ways to do so. After describing the interaction effect between the two problems, it suggests three pillars for protecting patient privacy while permitting data to be used for algorithmic verification. It concludes with a short case study of a related debate about access to information: the debate over access to clinical-trial information.

A. *Patient Privacy Versus Algorithmic Accountability*

The twin goals of algorithmic accountability and patient privacy are fundamentally at odds in black-box medicine, such that efforts to make one better will, in most instances, make the other worse. The close relationship between the privacy and accountability challenges arises because both stem from the massive datasets needed for black-box medicine to work.

Verifying black-box medical algorithms requires giving third parties access to large amounts of health information about thousands or millions of patients. Large amounts of data are needed because black-box algorithms and the human body are so complex that false positives and dead ends are

91. *Id.* at 10.

92. *See* 410 Ill. Comp. Stat. 50/3 & 50/4 (imposing a duty of confidentiality on physicians and providing civil penalties for the disclosure of confidential information).

common, so you need a lot of data to be confident that correlations are legitimate. And providing that data to third parties helps avoid conflicts of interest and ensure accountability in the verification process.[93] A verification process performed by a pharmaceutical company offering a new diagnostic test, for instance, would face different questions than one performed by the FDA or an independent academic researcher, since relying on the company that would benefit from the test creates a conflict of interest that would not exist with an independent verification.[94]

This conflict of interest is not new. Clinical trials for new diagnostic tests or devices have long been typically conducted by developers of those products, who have interests in seeing their products approved for market. Though several mechanisms are used to combat these conflicts of interest, these mechanisms would likely prove ineffective with black-box medicine. The FDA subjects clinical trials to strict rules,[95] including a requirement that all clinical trial materials—not just summaries or positive results—be submitted for independent review,[96] and another that all clinical trials be registered to avoid the suppression of negative results.[97] There have also been recent calls for greater disclosure of clinical-trial data to non-regulators, to permit more independent third-party analysis.[98] These measures are typically undergirded by scientific understanding of a drug or device's mechanism, which help ensure that it is an effective treatment. These mechanisms to combat conflicts of interest, though, don't work with black-box medicine; first-principles scientific understanding is unavailable because of the opaque

93. *See supra* Part II.B.

94. Further conflicts of interest can arise when the investigators conducting the trials have their own financial interests in the success of the company or the product. *See* Peter Whoriskey, *As drug industry's influence over research grows, so does the potential for bias*, WASHINGTON POST (Nov. 24, 2012), https://www.washingtonpost.com/business/economy/as-drug-industrys-influence-over-research-grows-so-does-the-potential-for-bias/2012/11/24/bb64d596-1264-11e2-be82-c3411b7680a9_story.html.

95. *See Selected FDA GCP/Clinical Trial Guidance Documents*, FDA, http://www.fda.gov/ScienceResearch/SpecialTopics/RunningClinicalTrials/GuidancesInformationSheetsandNotices/ucm219433.htm (last updated Aug. 12, 2016) (listing fifty guidance documents for sponsors of clinical trials).

96. *See* 21 C.F.R. § 314.50 (2015) (describing in detail the required content for a New Drug Application).

97. *See* 42 U.S.C. § 282(j) (2012) (establishing a "clinical trials registration data bank," later implemented as ClinicalTrials.gov). This requirement, though, is frequently flouted. *See* Charles Piller, *Law Ignored, Patients at Risk*, STAT NEWS (Dec. 13, 2015), http://www.statnews.com/2015/12/13/investigation/ (finding that among all institutions conducting 20 or more clinical trials since 2008, only two companies were at least 50% compliant with requirements to report results within one year of study completion or termination).

98. *See, e.g.*, INST. OF MED., SHARING CLINICAL TRIAL DATA: MAXIMIZING BENEFITS, MINIMIZING RISK (2015), *available at* http://nap.edu/18998 [hereinafter IOM, *Sharing Clinical Trial Data*]; Richard Lehman & Elizabeth Loder, *Missing Clinical Trial Data*, 344 BRIT. MED. J. d8158 (2012); Mary Beth Hamel et al., *Preparing for Responsible Sharing of Clinical Trial Data*, 369 N. ENGL. J. MED. 1651 (2013).

nature of black-box algorithms, and clinical trials are too expensive and too limited to be used for most algorithms. Third-party independent validation, then, is a necessary tool to overcome companies' conflicts of interest and ensure the quality and accuracy of black-box algorithms, and this means that large amounts of patient data must be made available to third parties.

Patient privacy works in the reverse direction, since both of these requirements for accountability—the large amount of information and the need to disclose it to third parties—lead to greater privacy problems. The more information that's collected, disclosed, and used, the greater the raw material for future privacy problems and the more serious those problems are likely to be when they do occur. If a doctor's office or hospital system, for instance, collects genetic profiles or detailed sexual histories for its patients, then a data breach is likely to be far more serious and lead to far greater privacy losses than if it just collected blood pressure and cholesterol readings. And the more often information is disclosed to third parties, the more likely privacy problems are to occur in the first place. This is the case both because the number of people who can cause a privacy problem is greater and because disclosures themselves—whether of paper or electronic media or by internet transmission—provide opportunities for data breaches and similar privacy problems. And the use of patient information can also lead to privacy problems because it results in the creation of new health information, which further exacerbates the privacy problems.

Because this relationship between accountability and privacy arises from the fundamental big-data nature of black-box medicine, severing the link between the two is unlikely to work. The trick, then, is to identify the best ways to balance the two interests. The next subpart discusses ways of doing so.

B. *Three Pillars for Privacy-Preserving Accountability*

Accountability and privacy in black-box medicine may be structurally opposed, but that doesn't mean that every effort to improve one of those values will affect the other to the same extent. Some efforts to promote accountability may destroy patient privacy, while other efforts may have only incidental effects on privacy. Likewise, some efforts to protect patient privacy will prove greater obstacles to accountability than other efforts.

To foster adoption by doctors and patients, reimbursement by insurers, and approval by regulators, developers of black-box algorithms must demonstrate their quality and reliability.[99] As described above, third-party verification is likely to be a crucial part of this process. The question is how companies can validate black-box algorithms without destroying patient privacy. This subsection identifies three pillars of a framework that policymak-

99. *See supra* Part II.A, II.C.

ers should use to preserve patient privacy with minimal downside for accountability, and vice-versa.

Substantive restrictions on data collection, use, and disclosure. The first pillar for protecting privacy while promoting accountability is a system of specific restrictions on the collection, use, and disclosure of patient data, so that those behaviors are not left to the free market. There are two reasons for this. One is that the privacy interests at stake are so great, given the potential privacy harms from the collection, use, or disclosure of patient health information. And the other is that companies' incentives are badly misaligned with those privacy interests. There is little incentive not to gather as much data as possible, whether or not it is useful for providing health care or developing and verifying black-box algorithms. Electronic storage is cheap, and even the risk of liability due to a data breach is minimized by court rulings making it difficult to bring data-breach lawsuits.[100] In industries like telecommunications and online services, companies are incented to purge information periodically to reduce the burden of responding to subpoenas and law-enforcement requests, but those burdens are minimal in the health-care industry. And once data is gathered, even for legitimate purposes, companies can benefit from using that data for other, borderline or illegitimate purposes, like marketing, discrimination, or even sale to data brokers.

The collection and use of patient health information present the most difficult regulatory challenges, because black-box medicine needs large amounts of health information to work. The whole point of black-box algorithms is that it is impossible to know, ex ante, what factors will be correlated with medical risks, effective treatments, or patient outcomes. This makes it critical to collect large amounts of medical information. At the same time, this information reveals so much about individual patients that large datasets could be easily abused. And nonmedical information, like marketing and billing information, likely plays no legitimate role in developing and verifying black-box algorithms.

Regulations that govern the collection and use of information for developing and verifying black-box algorithms, then, should permit broad collection of medical information. At the same time, they should require that information to be segregated from information used for non-black-box purposes and should prohibit its transfer for non-black-box uses. There are different ways to do this, but one simple method would be to require companies

100. *See, e.g.*, Clapper v. Amnesty Int'l., 133 S. Ct. 1138 (2013) (holding that federal courts lacked Article III standing based on "hypothetical future harm that is not certainly impending"); In re Science Applications Int'l Corp. Backup Tape Data Theft Litigation, 45 F. Supp. 3d 14 (D.D.C. 2014) (holding that the increased risk of identity theft after information about customers was stolen does not give rise to an injury in fact supporting Article III standing); *but see* Krottner v. Starbucks Corp, 628 F.3d 1139 (9th Cir. 2010) (finding Article III standing when plaintiffs "alleged a credible threat of real and immediate harm stemming from the theft of a laptop containing their unencrypted personal data").

developing black-box algorithms to use segregated write-only data vaults that can freely accept new data and be used to develop or verify algorithms but cannot export data for other uses.[101] Collection, use, and disclosure of data outside of these data vaults would continue to be governed by HIPAA and other laws. Because the data vaults would contain so much health information, they would be subject to heightened black-box regulations and would be the only data sources that companies and researchers could use to develop and verify black-box algorithms.

Regulations should also work to ensure that the data collected avoids bias due to selection effects. For instance, datasets should contain the same information about each patient, with the included data points selected according to neutral criteria, rather than just incorporating whatever data one can throw at the problem. Otherwise, there is a risk that differences in health information will introduce bias that could affect the resulting black-box algorithms. For instance, researchers have documented differences in physicians' pain-management treatment of black and white patients.[102] Similar differences in treatments, diagnostic tests ordered, physician notes, or even patients' propensity to seek medical treatment, whether from discrimination or from any other cause, could lead to systematic distortions of black-box algorithms.[103] The best way to counter these distortions is to rely on datasets that systematically capture the same information for each member of a representative group. The Precision Medicine Initiative and Million Veteran Program are valuable steps in this direction.[104]

Regulations of data disclosures present their own difficulties. The simplest, and least controversial, regulations would forbid using data for purposes beyond those commensurate with the purposes for which the data was

101. Write-only data vaults are not, though, a panacea. First, they are difficult to implement technologically, so their write-only nature would likely need to be enforced by data-use policies. The temptation to use patient health information for other purposes, though, like solving crimes, determining paternity, or even marketing new drugs, could be too great for policy-makers and companies developing black-box algorithms to withstand. And even if regulations hold, it is sometimes possible to reverse-engineer data from the outcome of an algorithm. One study, for instance, demonstrated that a machine-learning algorithm predicting optimal dosage of the drug warfarin could be inverted to predict patient genotypes based on the predicted dosage. *See* Matthew Fredrikson et al., *Privacy in Pharmacogenetics: An End-to-End Case Study of Personalized Warfarin Dosing*, *in* PROCEEDINGS OF THE 23RD USENIX SECURITY SYMPOSIUM 17 (2014). Truly opaque black-box algorithms would resist such reverse engineering, but it can be difficult to know if an algorithm is truly opaque, or if it just hasn't been successfully reverse engineered yet.

102. *See* Kelly M. Hoffman et al., *Racial Bias in Pain Assessment and Treatment Recommendations, and False Beliefs about Biological Differences between Blacks and Whites*, 113 PROC. NAT'L. ACAD. OF SCIS. 4296 (2016) (finding that African-Americans are systematically undertreated for pain relative to white Americans, due to false beliefs about biological differences between blacks and whites).

103. *See* Anupam Chander, *The Racist Algorithm?,* 115 MICH. L. REV. (forthcoming 2017).

104. *See supra* notes 30-31 and accompanying text.

collected. Data that has been used to develop and verify black-box algorithms, then, could be used to verify those algorithms or, perhaps, to develop new algorithms, but couldn't be used for marketing. But restrictions of this kind may not do enough to protect privacy, since detailed patient-specific information could still wind up in numerous hands, with regulations and contractual restrictions acting as imperfect restraints on behavior. And because this data presents so many privacy issues—deidentification is essentially impossible with the sorts of granular genetic data used to develop black-box algorithms[105]—imperfect restraints could quickly turn into no restraints.

A stronger approach would be to tailor the degree of access parties have to their needs and to encourage data collectors to make data available in more granular forms. Much work developing and verifying algorithms, for instance, does not require handing over raw data, but could work perfectly well with access to query-and-response systems.[106] Using these systems, third parties developing algorithms could submit test algorithms to an interface offered by a data collector and receive the output from executing those algorithms on real patient data, without providing those third parties access to that patient data.[107] Or, third parties testing algorithms on new data could submit that data to a black-box algorithm and receive the results without access to the details of the underlying algorithm. Such query-and-response systems would not work all the time, but when they do, they can help minimize the amount of patient information transmitted and disclosed to third

105. In other contexts, experts have recommended deidentification of patient data to protect privacy. We do not do so here, for two reasons. First, the power of black-box medicine comes from its ability to find patterns in large amounts of patient data. It is impossible to know ex ante, then, what information can be removed without degrading the resulting black-box algorithm, with trivial exceptions like name and zip code. Accordingly, deidentification risks producing substantively worse health-care outcomes. And second, reidentification is a significant-enough risk with any deidentification scheme that deidentification can act as a false security blanket, reassuring individuals that privacy risks are lower than they are. *See generally* INST. OF MED., *supra* note 74.

106. The same can be true of verification. *See, e.g.,* Philip Adler et al., *Auditing Black-box Models by Obscuring Features,* http://arxiv.org/abs/1602.07043 (Feb. 23, 2016).

107. Such an approach was used for the Netflix Prize, a contest in which Netflix offered independent developers a $1 million prize for improving its movie-recommendation algorithm. Netflix made a training dataset available to developers and let them test algorithms against a larger dataset of customer ratings, which was not otherwise available to developers. *E.g.,* Jason Kincaid, *The Netflix Prize Comes To A Buzzer-Beater, Nailbiting Finish,* TechCrunch, https://techcrunch.com/2009/07/26/the-netflix-prize-comes-to-a-buzzer-beater-nailbiting-finish/ (Jul. 26, 2009). The Netflix Prize story has a cautionary coda, though: Even the training dataset had enough data to identify several individuals, when combined with another dataset like IMDB's review database. *See* Arvind Narayanan & Vitaly Shmatikov, *Robust Deanonymization of Large Sparse Datasets,* in Proc. of 29th IEEE Symposium on Security and Privacy 111-125 (2008); Ryan Siegel, *Netflix Spilled Your* Brokeback Mountain *Secret, Lawsuit Claims,* Wired, https://www.wired.com/2009/12/netflix-privacy-lawsuit/ (Dec. 17, 2009).

parties. And when they don't work, there are also data-science techniques that can minimize third parties' *access* to sensitive patient data while still providing the ability to *use* that data. Homomorphic-encryption systems, for instance, allow analysis to be done on encrypted data, giving encrypted results that can then be decrypted without providing access to the decrypted data. These schemes are relatively immature, but they have been used to permit confidentiality-preserving big-data analysis.[108]

Independent gatekeepers governing access to patient data and black-box models. The second pillar is a system of independent data gatekeepers to determine when, and under what conditions, researchers can get access to patient data and black-box medical models. These independent gatekeepers could be government entities located in agencies like the Food and Drug Administration or National Institutes of Health or could be parts of international or nongovernmental organizations like the World Health Organization. Regardless, the key is that companies that develop black-box algorithms not have free rein to decide what patient data to collect, use, and disseminate, and how to do so, since those companies' incentives are badly misaligned with patient privacy and since the privacy interests at stake are so great. Instead, independent assessment of a plan for data-sharing would help ensure that the privacy interests of patients are considered, not just private commercial considerations.

There are different ways to design a system of independent gatekeepers. Gatekeepers could have authority to require companies to collect, use, or disseminate specific data in specific ways or could simply authorize voluntary actions by companies. A comprehensive and mandatory system, like the FDA's drug-approval process, could pair regulatory approval of products based on black-box algorithms with requirements for what data collection must take place, how that data must be stored and used, what verification steps must occur, and who must perform that verification. Such a system could require premarket approval, could condition approval of a black-box model on independent verification within a set time frame, or could use market mechanisms or other non-mandate incentives to encourage verification. Or, a system tailored to protecting privacy without comprehensively regulating algorithmic medical products, like the HIPAA privacy rule or any of a variety of other domain-specific privacy laws, could place limits on companies' collection, use, and disclosure of data while leaving algorithmic verification to the FDA or someone else.

Either way, companies should be required to obtain advance permission from the independent gatekeeper before collecting, disclosing, and using pa-

---

108. *See* Julian James Stephen et al., *Practical Confidentiality Preserving Big Data Analysis, in* 6TH USENIX WORKSHOP ON HOT TOPICS IN CLOUD COMPUTING (2014), *available at* https://www.usenix.org/system/files/conference/hotcloud14/hotcloud14-stephen.pdf (providing a proof of concept of the "ability to maintain sensitive data only in an encrypted form in the cloud and still perform meaningful data analysis").

tient information in the ways most likely to lead to privacy losses.[109] This means, at least, before making specific uses of patient data or disclosing that data to third parties. And before granting that permission, an independent gatekeeper should balance patient privacy interests with the need for verification and the specific techniques of data collection, storage, use, and disclosure that a company proposes to use. A company should not hand over raw patient data, for instance, if a query-and-response system would accomplish the same goal.

Information-security requirements. The third pillar is a system of security requirements governing the storage and transmission of medical information. While the first two pillars are designed to avoid privacy losses due to the actions of researchers developing or verifying black-box models, this pillar is designed to avoid losses due to third parties. As near-daily reports of data breaches make clear, the most acute threats to privacy may not be the actions of those with legitimate access to personal information, but instead the actions of criminals who obtain that information illegally. And the same may be true in black-box medicine: If researchers developing or verifying black-box models maintain or transfer the underlying data insecurely, then it doesn't much matter what those with legitimate access do because criminals will soon have free rein to use patient information as they wish.

A detailed guide to modern security practices is beyond the scope of this article,[110] but there are several basic practices that should be included in any security plan involving sensitive information like patient health data. That information should be protected with strong encryption in both storage and transmission. Access should be limited to individuals with legitimate needs and should be person-specific so access can be monitored and revoked. Systems should use two-factor authentication instead of easy-to-guess pass-

109. Such permission could act as a reasonable substitute for consent from each patient or information subject. Requiring individual consent from a patient before health information is included in a dataset from which black-box algorithms are developed could, in some circumstances, impose significant obstacles to algorithm development. *Cf. Clinical Research and the HIPAA Privacy Rule*, NAT'L INSTS. OF HEALTH (Feb. 2004), https://privacyruleandresearch.nih.gov/pdf/clin_research.pdf. Relying on patient consent can also create bias in the dataset, as patients willing to consent differ from those unwilling to consent. *See* INSTITUTE OF MEDICINE, BEYOND THE HIPAA PRIVACY RULE: ENHANCING PRIVACY, IMPROVING HEALTH THROUGH RESEARCH, 209–14 (2009). In some cases, though, as in the case of comprehensive volunteer datasets assembled for no reason other than algorithm development, individual consent may be more appropriate or even necessary.

110. A good place to start would be several standards from the International Organization for Standardization and the International Electrotechnical Commission governing privacy and security of data stored in cloud systems. *See* ISO/IEC 27001: INFORMATION SECURITY MANAGEMENT SYSTEMS - REQUIREMENTS (2013); ISO/IEC 27002: CODE OF PRACTICE FOR INFORMATION SECURITY CONTROLS (2013); ISO/IEC 27017: CODE OF PRACTICE FOR INFORMATION SECURITY CONTROLS BASED ON ISO/IEC 27002 FOR CLOUD SERVICES (2015); ISO/IEC 27018: CODE OF PRACTICE FOR PROTECTION OF PERSONALLY IDENTIFIABLE INFORMATION (PII) IN PUBLIC CLOUDS ACTING AS PII PROCESSORS (2014).

words. Access logs should be kept and routinely monitored for unusual access patterns. Individuals with access should receive security training, including training on social engineering and other factors that lead to security breaches. That training should be refreshed periodically.

Most critically, these security practices must be continuously updated, since tools and techniques that are sufficient when adopted can be out of date weeks or months later, as security vulnerabilities are discovered. This need for constant updating presents problems for regulators, since rules setting forth specific requirements can only be updated so often. One solution has been to rely on industry standards, which can evolve over time, rather than detailed regulations. In lieu of specific security regulations, for instance, the Federal Trade Commission has relied on quasi-common-law enforcement of reasonable security standards, an approach the Third Circuit blessed in 2015.[111] Such an approach might work with black-box medicine, though it would be harder than in other contexts, since the harm from a breach would be greater. Rather than rely on reasonable industry-standard practices, then, a rule would need to require something like best practices, which are harder and more expensive to maintain. Or, a rule could rely on large post-breach penalties to persuade companies to use strong security—a risky approach, since security risks are easy to underestimate and the costs of a breach would be so high.

None of these three pillars is unique; all three represent elements of existing privacy-protection schemes like the Fair Information Practices[112] or the European Union's Data Protection Directive.[113] The first pillar, substantive limitations on data collection, use, and disclosure, reflects principles like "respect for context" and "focused collection" enumerated in the Obama administration's proposed Consumer Bill of Rights. That proposal suggests

---

111. *See* F.T.C. v. Wyndham Worldwide Corp., 799 F.3d 236 (3d Cir. 2015); *see also* Daniel J. Solove & Woodrow Hartzog, *The FTC and the New Common Law of Privacy*, 114 COLUM. L. REV. 583 (2014) (arguing that in *Wyndham*, the "defendant's arguments against the FTC's detailed security requirements neglect[ed] to acknowledge that FTC jurisprudence has progressed in a natural and logical fashion. One would expect over time for a general standard about data security to be refined as that standard is applied in specific cases. This is an almost inevitable progression, and it is exactly how the common law works.").

112. The FIPs, sometimes called the FIPPs (for Fair Information Practice Principles), have a long and convoluted history in the law. The FIPs originated in the 1970s in the Department of Health, Education & Welfare, and have since been articulated in various forms in American law by the Federal Trade Commission; the National Science and Technology Council; and the Departments of Homeland Security, Commerce, and Health and Human Services, and internationally by the Organization for Economic Cooperation and Development. For an overview of this history, *see generally* Robert Gellman, *Fair Information Practices: A Basic History* (June 17, 2016), http://bobgellman.com/rg-docs/rg-FIPShistory.pdf; Paul M. Schwartz, *Preemption and Privacy*, 118 YALE L.J. 902, 907–08 (2009); PRISCILLA M. REGAN, LEGISLATING PRIVACY: TECHNOLOGY, SOCIAL VALUES, AND PUBLIC POLICY 73–86 (1995).

113. Council Directive 95/46, 1995 O.J. (L 281) 31 (EC). [hereinafter Data Protection Directive].

that companies should limit their data collection, use, and disclosure so they are consistent with the contexts in which consumers originally disclosed data.[114] Likewise, the Data Protection Directive imposes proportionality and legitimate-purpose requirements that limit the purposes for which personal information may be processed.[115] The independent-gatekeeper model is reminiscent of the EU's data protection authorities, which have substantive authority over the processing of personal information.[116] And both the Data Protection Directive and several implementations of the FIPs require custodians to take steps to keep data secure.[117]

The approach we describe in this section is both narrower and broader than the principles embraced in the FIPs and the Data Protection Directive, since the context of black-box medicine presents specific privacy needs that do not apply universally.[118] The approach is narrower because unlike both the FIPs and the Data Protection Directive, we do not focus on giving information subjects rights to notice of, consent to, or control over the uses to which information is put. Such individual rights are the most prominent piece of most formulations of the FIPS and of privacy law in the United States generally; they also represent an important piece of the EU's data-protection system.[119] And they may be appropriate in the case of black-box medicine, for reasons stemming from general privacy principles, or in certain contexts that arise in the development of black-box medicine. But a

---

114. WHITE HOUSE, CONSUMER DATA PRIVACY IN A NETWORKED WORLD: A FRAMEWORK FOR PROTECTING PRIVACY AND PROMOTING INNOVATION IN THE GLOBAL DIGITAL ECONOMY 15, 21 (2012), *available at* https://www.whitehouse.gov/sites/default/files/privacy-final.pdf. Not all formulations of the FIPs impose similar requirements. *See infra* notes 115–119 and accompanying text.

115. Data Protection Directive, *supra* note 113, at arts. 6-7. *See also* Commission Regulation 2016/679, 2016 O.J. arts. 4-6 (L 119) 1 [hereinafter General Data Protection Regulation] (superseding the Data Protection Directive in 2018 with similar provisions).

116. *See* Data Protection Directive, *supra* note 113, at art. 28; General Data Protection Regulation, *supra* note 115, at arts. 51-59.

117. Data Protection Directive, *supra* note 113, at art. 17. *See also* FEDERAL TRADE COMMISSION, PRIVACY ONLINE: A REPORT TO CONGRESS 10 (1998), *available at* https://www.ftc.gov/sites/default/files/documents/reports/privacy-online-report-congress/priv-23a.pdf (describing "integrity/security" as one of five widely accepted principles of privacy protection); WHITE HOUSE, *supra* note 114, at 19 (stating that consumers "have a right to secure and responsible handling of personal data").

118. *Compare, e.g.,* James G. Hodge, Jr., et al., *Legal Issues Concerning Electronic Health Information: Privacy, Quality, and Liability,* 282 J. AM. MED. ASSN. 1466 (1999).

119. *See, e.g.*, Daniel J. Solove, *Privacy Self-Management and the Consent Dilemma*, 126 HARV. L. REV. 1880, 1880 (2013) ("[T]he basic approach to protecting privacy has remained largely unchanged since the 1970s. . . . The law provides people with a set of rights to enable them to make decisions about how to manage their data. . . . I will refer to this approach to privacy regulation as 'privacy self-management.' "); White House, *supra* note 114, at 11-15 (asserting that consumers have rights "to exercise control over what personal data companies collect from them and how they use it" and "to easily understandable and accessible information about privacy and security practices"); Data Protection Directive, *supra* note 113, at arts. 9-12, 14-15.

notice-and-consent regime would make it much harder to obtain the large amounts of comprehensive health information needed for black-box medicine to work. Since the potential social value of black-box medicine is so large, tools like the substantive limitations described above would better serve the twin goals of protecting patient privacy and encouraging the development of high-quality black-box algorithms. And these tools are broader because the substantive limitations on collection, use, and distribution of health information and the duties of an independent gatekeeper are far more detailed and limiting than the general principles enumerated in the FIPs and the Data Protection Directive. Because the potential privacy harms from misuse of health information are so great, safeguards not normally used in other industries, like advance gatekeeper approval before a company engages in any data collection, use, or distribution, are likely necessary.

Each of the three pillars discussed in this section is an important component of any scheme to encourage verification of black-box medicine while preserving patient privacy because each addresses a different threat model. Substantive restrictions on data collection, use, and disclosure help ensure that developers of black-box algorithms, who face strong incentives to disregard patient privacy, are restrained from doing so. Independent gatekeepers help ensure that patient data and black-box models are shared with researchers who can perform independent verification while removing conflicts of interest and ensuring that that sharing avoids posing greater risks to patient privacy. And information-security requirements help ensure that malicious outsiders cannot get hold of patient information, preventing data breaches. Without each component, key openings would remain for patient privacy to be lost without any corresponding benefit.

C. *Case Study: Data and the Clinical-Trial Debate*

Experience with similar data-sharing problems in other health-care contexts shows the importance of the three pillars discussed in the last section. Clinical-trial data sharing, for instance, provides a useful case study of how privacy can be built into a system in which detailed medical information is shared between companies, researchers, and the government.[120] When researchers carry out clinical trials to test new drugs or other medical interventions, they collect large amounts of data about patients, including how those patients respond to the new drug or intervention.[121] Some of these data are analyzed to determine the results of the study; some of those results are

120. *See* IOM, *Sharing Clinical Trial Data*, *supra* note 98 (explaining that the committee "analyzes how several risks associated with sharing clinical trial data (in particular individual participant data and CSRs) might be addressed through controls on data access (i.e., with whom the data are shared and under what conditions) without compromising the usefulness of data sharing for the generation of additional scientific knowledge").

121. *Id.* at 18 ("Vast amounts of data are generated over the course of a clinical trial.").

published, though many are not.[122] Whether or not a study is published, however, much of the data collected is never analyzed; and almost all of these data are typically kept proprietary.[123] Although such proprietary data are disclosed to the FDA when a study's sponsor seeks regulatory approval for a new drug or intervention, only study summaries and results are usually available to others.

This paradigm of limited data sharing has been challenged in recent years by researchers and other third parties seeking broader disclosure of raw data from clinical trials.[124] These third parties have different goals. Some seek to use raw clinical-trial data to validate published results, combining data from different trials into larger datasets and performing secondary analyses.[125] Others seek to speed drug discovery, find new drug targets, or identify intermediate clinical goals that are easier to measure.[126] All of these goals, like the goals of black-box medicine, seek to find new patterns in previously disparate sources of health information. And just as in black-box medicine, clinical-trial data sharing faces interrelated problems of accountability and privacy: Sharing clinical-trial data helps third parties validate the results of clinical trials, while exposing health information in ways that could lead to privacy harms.

In the wake of these increasing calls for clinical-trial data sharing, scholars and policymakers have considered numerous regulatory questions, including how to make data available for independent analysis and verification without sacrificing patient privacy.[127] The most prominent effort culminated

---

122. *See* Peter Doshi et al., *Restoring Invisible and Abandoned Trials: A Call for People to Publish the Findings,* 346 BMJ f2865 (2013) (observing that one "basic proble[m] of representation driving growing concerns about relying on published research to reflect truth . . . is no representation (invisibility), which occurs when a trial remains unpublished years after completion.").

123. *Id.*

124. *See, e.g.*, Peter Doshi et al., *Raw Data from Clinical Trials: Within Reach?*, 34 TRENDS PHARMACOLOGICAL SCI. 645 (2013) ("Making raw data from clinical trials widely publically available should reduce selective reporting biases and enhance the reproducibility of and trust in clinical research . . . [though] the optimal procedures for data sharing are hotly debated."); Hamel et al., supra note 98 ("Data from clinical trials, including participant level data, are being shared by sponsors and investigators more widely than ever before. . . . [It] may bring exciting benefits for scientific research and public health but may also have unintended consequences. Thus, expanded data sharing must be pursued thoughtfully.").

125. IOM, *Sharing Clinical Trial Data*, *supra* note 98, at 18 ("Today, researchers other than the trialists have limited access to clinical trial data that could be used to reproduce published results, carry out secondary analyses, or combine data from different trials in systematic reviews. Public well-being would be enhanced by the additional knowledge that could be gained from these analyses.").

126. *Id.*

127. Other questions of clinical-trial data governance, including avoiding data misuse, protecting intellectual property, preventing undue commercial harms to those sharing data, and enhancing public trust, may also have relevance for black-box medicine but are beyond the scope of this article. *See id.* at 139–58.

in a report from the Institute of Medicine in 2015.[128] The report called for open access to the results of clinical trials, but more limited access to raw data from those trials, reasoning that such a compromise would best balance the legitimate interest in data sharing with the risks, burdens, and challenges it would bring.

Critically, in making recommendations for ways to protect privacy, the Institute of Medicine report included elements of each of the pillars discussed in the last section. On the first pillar (substantive restrictions on the collection, use, and disclosure of data), it made relatively few recommendations, since the information collected in clinical trials is otherwise regulated by the FDA. But it made several recommendations designed to minimize the risk of data disclosures. It recommended deidentification, as is typical in health research,[129] while noting the dangers of reidentification and recommending that recipients of data should commit not to intentionally reidentify data subjects.[130] And although clinical-trial data contains large amounts of detailed patient health information, the report recommended against more robust deidentification techniques that degrade data quality by, for instance, removing details.[131] Since these techniques would harm the utility of the data, the report instead called for combining surface-level deidentification methods with information-security measures—the third pillar discussed above—to safeguard data against inadvertent and unauthorized access.[132]

The most detailed recommendations in the Institute of Medicine report concerned procedural safeguards regulating access to clinical-trial data, including an independent gatekeeper like the one discussed in the second pillar above. The report recommended that access to clinical-trial data be moderated by a gatekeeper, ideally an independent panel that could evaluate the expertise and research objectives of entities seeking access.[133] It also recommended that that gatekeeper require those receiving data to commit to data-use agreements prohibiting unapproved uses such as reidentification, contact with clinical-trial participants, and further data sharing.[134] And it recommended that data access be transparent.[135]

Groups are working to implement the committee's recommendations. One prominent group acting as an independent gatekeeper is the Yale Open Data Access (YODA) Project, which partners with pharmaceutical compa-

128. *Id.*
129. *Id.* at 144.
130. *Id.* at 145-46.
131. *Id* at 146.
132. *Id.* at 146–47.
133. *Id.* at 149–56.
134. *Id.* at 147–48. The committee noted other common data use agreement provision, including assignments of intellectual property, prohibitions on competitive commercial use, acknowledgement requirements in publications, and restrictions on non-proposed data uses. *Id.*
135. *Id.* at 156.

nies to provide access to patient-level clinical-trial data.[136] As of July 2016, YODA has data from nearly 200 clinical trials, which it makes available to researchers using it to further science and public-health goals.[137] In addition to serving as an independent gatekeeper, YODA imposes policies that promote the other pillars of privacy-preserving information sharing, include requirements for transparency of access, data security, prior assessment of researcher and research quality, and data-use Agreements limiting further dissemination or commercial use of data.[138]

Projects like YODA show that the committee's recommendations and analyses provide helpful examples of how to implement a system of robust data sharing while protecting patient privacy—exactly what is needed for black-box medicine to thrive. Not all of its recommendations apply to black-box medicine. In particular, some technological means of protecting data, like distributed datasets and the introduction of random noise, would be problematic for developers of black-box medicine, who rely on large unified datasets (making distribution problematic) and often cannot predict what data will be useful (making it hard to introduce random noise without harming the utility of the datasets). However, the report's analysis reinforces the roles that the three basic pillars of privacy-preserving accountability can play in providing a useful framework for protecting privacy in black-box medicine.

Conclusion

Black-box medicine could transform health care, but to do so it must first overcome the twin challenges of algorithmic accountability and patient privacy. These challenges stem from the big-data nature of black-box medicine. Researchers need access to massive amounts of health information to develop black-box algorithms, putting patients at risk of privacy losses. And independent researchers need access to this same information to verify black-box algorithms, ensuring they are accurate and unbiased, but risking further privacy losses. Balancing this tension between accountability and privacy is a key challenge in the development of black-box medicine.

To best accommodate these competing challenges, policymakers should look to a framework of privacy-preserving accountability built on three pillars. First, researchers developing black-box algorithms should comply with substantive limitations on the collection, use, and disclosure of patient health information. Second, independent gatekeepers should oversee information sharing between those developing and verifying black-box algorithms. And

136. The YODA Project, http://yoda.yale.edu/ (last visited Nov. 6, 2016).

137. *Id.*; *Policies & Procedures to Guide External Investigator Access to Clinical Trial Data*, The YODA Project, http://yoda.yale.edu/policies-procedures-guide-external-investigator-access-clinical-trial-data (last visited Nov. 6, 2016).

138. *Id.*

third, robust information-security requirements should be imposed to prevent unintentional data breaches of patient information. By developing rules based on these three pillars, regulators can help ensure that patients obtain the substantial benefits of black-box medicine without sacrificing their privacy.