2016

Big Data, Patents, and the Future of Medicine

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

Recommended Citation

Price, W. Nicholson, II. "Big Data, Patents, and the Future of Medicine." *Cardozo Law Review* 37, no. 4 (2016): 1401-53.

BIG DATA, PATENTS, AND THE FUTURE OF MEDICINE

W. Nicholson Price II†

Big data has tremendous potential to improve health care. Unfortunately, intellectual property law isn't ready to support that leap. In the next wave of data-driven medicine, black-box medicine, researchers use sophisticated algorithms to examine huge troves of health data, finding complex, implicit relationships and making individualized assessments for patients. Black-box medicine offers potentially immense benefits, but also requires substantial high investment. Firms must develop new datasets, models, and validations, which are all nonrivalrous information goods with significant spillovers, requiring incentives for welfare-optimizing investment.

Current intellectual property law fails to provide adequate incentives for black-box medicine. The Supreme Court has sharply restricted patentable subject matter in the recent Prometheus, Myriad, *and* Alice *cases, and what might still be patentable is limited by the statutory requirements of written description and enablement. Other incentives for investment, such as trade secrecy or prizes, fail to fill the gaps. These limits push firms away from using big data in medicine to solve big problems, and push firms toward small-scale incremental innovation. Small tweaks to doctrine will help, but are not enough. Instead, the big data needed to support transformative medical innovation should be considered as infrastructure for innovation and should be the focus of substantial public effort.*

† Assistant Professor, University of New Hampshire School of Law. J.D., Columbia University School of Law, 2011; Ph.D. (Biological Sciences), Columbia University Graduate School of Arts and Sciences, 2010. I wish to thank Ana Bračič, Glenn Cohen, Rebecca Eisenberg, Roger Ford, Matt Lawrence, Michael Madison, Peter Menell, Kevin Outterson, Geertrui Van Overwalle, Ben Roin, Rachel Sachs, Jake Sherkow, and Jeff Skopek for their helpful comments and feedback. Cassandra Simmons and Abhishek Shukla-Bannerjee provided excellent research assistance. This work benefited from feedback at the Health Law Professors' Conference, the Munich Conference on Innovation and Competition, the Intellectual Property Scholars' Conference, the Michigan State Junior Intellectual Property Scholars' Workshop, the University of Tulsa Faculty Colloquium, the Suffolk IP Roundtable, and the Boston University Workshop on Personalized Medicine and Incentives. All errors are my own.

TABLE OF CONTENTS

INTRODUCTION

Medicine is both expensive and imprecise. Although doctors have an increasingly expansive toolkit of treatment options, knowing which drug to give to which patient—and how much and when—requires substantial knowledge about differences between patients and the biological networks that underlie treatments. Personalized medicine tries to clear the fog, determining the characteristics of each patient and her disease and recommending the most appropriate treatment.[1] Personalized medicine has been a scientific and policy goal for years, but has recently received renewed policy focus.[2] In President Obama's 2015 State of the Union Address, he announced the Precision Medicine Initiative,[3] aimed at driving research and development of personalized medicine.[4]

The dominant examples of personalized medicine so far are relationships that are well-understood and validated in clinical trials: For instance, using a single genetic test to find whether a patient's cancer is likely to respond to a drug developed alongside that test, and treating the patient accordingly.[5] But simple one-to-one relationships are only a relatively small part of biomedical complexity,[6] and it is much harder to fully understand and clinically validate more complex relationships.[7] Diseases and treatments are frequently dependent on combinations of multiple genetic variables with environmental factors

---

1 *See infra* text accompanying notes 32–40.

2 *See, e.g.*, Wylie Burke & Bruce M. Psaty, *Personalized Medicine in the Era of Genomics*, 298 J. AM. MED. ASS'N 1682 (2007); Isaac S. Chan & Geoffrey S. Ginsburg, *Personalized Medicine: Progress and Promise*, 12 GENOMICS & HUM. GENETICS, 2011, at 217; Geoffrey S. Ginsburg & Jeanette J. McCarthy, *Personalized Medicine: Revolutionizing Drug Discovery and Patient Care*, 19 TRENDS BIOTECHNOLOGY 491 (2001); Margaret A. Hamburg & Francis S. Collins, *The Path to Personalized Medicine*, 363 NEW ENG. J. MED. 301 (2010); Pauline C. Ng et al., *An Agenda for Personalized Medicine*, 461 NATURE 724 (2009).

3 President Barack H. Obama, Address Before a Joint Session of Congress on the State of the Union (Jan. 20, 2015) (transcript available at https://www.whitehouse.gov/the-press-office/2015/01/20/remarks-president-state-union-address-january-20-2015).

4 *See* Francis S. Collins & Harold Varmus, *A New Initiative on Precision Medicine*, 372 NEW ENG. J. MED. 793, 793 (2015). For the purposes of this Article, "personalized medicine" and "precision medicine" are used synonymously. The Precision Medicine Initiative is discussed in greater detail below. *See infra* notes 185–90.

5 *See* Walter P. Carney, *HER2/neu Status Is an Important Biomarker in Guiding Personalized HER2/neu Therapy*, 9 CONNECTION 25 (2006) (discussing the use of the drug Herceptin to treat breast cancer after clinical trials confirmed that Herceptin was effective only when the tumor overexpresses the gene for a particular receptor that the drug targets).

6 *See* Soumita Podder & Tapash C. Ghosh, *Exploring the Differences in Evolutionary Rates Between Monogenic and Polygenic Disease Genes in Human*, 27 MOLECULAR BIOLOGY EVOLUTION 934, 934 (2010) (noting that simple one-gene genetic disorders are much less common than multifactorial genetic diseases).

7 W. Nicholson Price II, *Black-Box Medicine*, 28 HARV. J.L. & TECH. 419, 441 (2015) [hereinafter Price, *Black-Box Medicine*].

and other physical variables, like weight, blood pressure, and sex. To find these dependencies and relationships, personalized medicine cannot rely on figuring out everything explicitly and confirming via clinical trials.[8] Instead, scientists can use sophisticated computer algorithms to analyze large datasets of health information, seeking patterns, predictions, and recommendations. This is black-box medicine.[9]

Black-box medicine is "black-box" precisely because the relationships at its heart are opaque—not because their developers deliberately hide them, but because either they are too complex to understand, or they are the product of non-transparent algorithms that never tell the scientists, "this is what we found."[10] Opacity is not desirable, but is rather a necessary byproduct of the development process.[11]

Black-box medicine lets scientists tap a wider range of biological relationships, and carries correspondingly broad benefits for health care. Matching patients to diseases and treatments more precisely could improve the quality of treatment, reduce the incidence of unnecessary side effects, and potentially save billions in wasted or inappropriate medical care.[12] It also suggests the possibility of new treatments, whether by suggesting new possibilities for drug exploration or by repurposing already-approved drugs for new or more targeted uses.[13] Rather than health care decisions being driven only by a relatively small set of carefully controlled clinical trials conducted on broad categories, decisions could be informed by the ongoing and collective medical experience of hundreds of millions of other patients. Black-box

---

8 *See* P.M. Rothwell, *Can Overall Results of Clinical Trials Be Applied to All Patients?*, 345 LANCET 1616 (1995).

9 *See* discussion *infra* Section I.A.

10 Price, *Black-Box Medicine*, *supra* note 7, at 433–34.

11 *Id.* at 434. This process-based opacity contrasts with situations where those developing information and algorithms deliberately keep them secret, whether for competitive advantage, to avoid public or government scrutiny, or for other reasons. For a description of the problems with deliberate secrecy and obscurity, see generally FRANK PASQUALE, THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION (2015).

12 PRESIDENT'S COUNCIL OF ADVISORS ON SCI. & TECH., PRIORITIES FOR PERSONALIZED MEDICINE 1 (2008), https://www.whitehouse.gov/files/documents/ostp/PCAST/pcast_report_v2.pdf.

13 Benjamin N. Roin, *Solving the Problem of New Uses*, 11 WASH. J.L. TECH. & ARTS (forthcoming 2016) [hereinafter Roin, *Solving the Problem of New Uses*] ("The vast majority of drug compounds operate by targeting biological pathways that may affect the progress or symptoms of a range of diseases, and almost all drugs have 'off-target' activity on other biological pathways that may affect a different set of diseases. Consequently, it is common that a drug designed to treat one disease will have potential new indications for treating one or more entirely different conditions." (footnotes omitted)).

medicine promises to radically expand the reach of personalized medicine, with tremendous potential gains.[14]

The question, then, is how to get there, and the path is not straightforward. Costs and hurdles exist at each phase of black-box medicine's development.[15] First, information must be gathered and vetted, which requires financial resources and navigating legal requirements, including privacy and informed consent.[16] Second, developing reliable and sensitive algorithms demands dedicated efforts by sophisticated programmers.[17] The experience of other predictive algorithms demonstrates this; for example, the movie-rental service Netflix created a three-year, multi-million dollar prize effort to improve its simple movie-prediction algorithm, in which thousands of teams managed to improve the algorithm's performance by only ten percent.[18] Third, since complex implicit predictions are much less amenable to the forms of validation on which we traditionally rely—scientific understanding, clinical trials, and postmarket surveillance—other forms of validation must be developed by the innovating firm, regulators, or third parties—or some combination of the three.[19]

Overcoming these hurdles will require significant incentives, and pure market incentives are likely to be woefully insufficient. Black-box medicine follows the classic pattern justifying intellectual property, in which firms underinvest in non-excludable information goods because they cannot capture the full social value of those goods.[20] Black-box medicine relies principally on pure information goods: collected data, patterns discovered within that data, and validation of those patterns.[21] Intellectual property protection theoretically allows firms to exclude others from the information good and therefore appropriate a higher portion—though not all—of the surplus, increasing innovation closer to optimal levels.

---

14 *See* discussion *infra* Section I.A.

15 *See* discussion *infra* Section I.B.

16 *See* discussion *infra* Section I.B.1.

17 *See* discussion *infra* Section I.B.2.

18 *See* Price, *Black-Box Medicine*, *supra* note 7, at 439 (describing the process of Netflix updating its movie recommendation algorithm); *The Netflix Prize Rules*, NETFLIX, http://www.netflixprize.com/rules (last visited May 9, 2015); Prizemaster, NETFLIX PRIZE (Sept. 18, 2009 4:58 PM), http://www.netflixprize.com/community/viewtopic.php?id=1537 (announcing the winner and noting a 10.06% improvement in performance).

19 *See* discussion *infra* Section I.B.3.

20 *See, e.g.*, Kenneth J. Arrow, *Economic Welfare and the Allocation of Resources for Invention*, *in* THE RATE AND DIRECTION OF INVENTIVE ACTIVITY: ECONOMIC AND SOCIAL FACTORS 609, 619 (1962) ("To sum up, we expect a free enterprise economy to underinvest in invention and research (as compared with an ideal) because it is risky, because the product can be appropriated only to a limited extent, and because of increasing returns in use.").

21 *See* discussion *infra* Section I.A.

The current intellectual property regime not only provides inadequate incentives for black-box medicine, but the incentives it provides also push the field in counterproductive directions. Patents provide the primary intellectual property incentives for technological innovation, and although patents are imperfect at driving algorithm development, they still create significant incentives.[22] Until quite recently, method patents were broadly available for diagnostic algorithms, as long as they satisfied the Federal Circuit's requirement that the invention involve a machine or a transformation of matter—which could be satisfied by as little as performing a blood test.[23] But in 2012, the Supreme Court held in *Mayo Collaborative Services v. Prometheus Laboratories* that a patent covering a standard diagnostic method—administering a drug, measuring the level of a metabolite, and knowing based on the result whether to increase or decrease the drug's dosage—was unpatentable, as essentially claiming, and thus preempting, a law of nature.[24] Close on the heels of *Prometheus*, the Court decided *Association for Molecular Pathology v. Myriad Genetics* in 2013, holding that isolated genomic DNA is unpatentable as a natural phenomenon;[25] such DNA patents, while not essential to diagnostic testing methods, provided secondary protection to those methods involving genetic testing.[26] Finally, in 2014, the Court in *Alice Corp. v. CLS Bank International* strengthened *Prometheus* by holding that abstract inventions, such as an algorithm, were not made patentable merely by implementing them on a computer.[27]

---

22 *See infra* Section II.A.

23 When the Federal Circuit first addressed the *Prometheus* case, it held that testing blood for the presence of metabolites was a "transformation" sufficient to make the invention patentable. Prometheus Labs., Inc. v. Mayo Collaborative Servs., 581 F.3d 1336, 1347 (Fed. Cir. 2009), *vacated*, 561 U.S. 1040 (2010). In the Supreme Court's decision in *Bilski v. Kappos*, the Federal Circuit's "machine or transformation test" went from a dispositive test to an "important and useful clue" as to whether the invention covers patentable subject matter. Bilski v. Kappos, 561 U.S. 593, 602–04 (2010). However, the importance of this clue to the Federal Circuit was such that it remained practically dispositive. *See* Prometheus Labs., Inc. v. Mayo Collaborative Servs., 628 F.3d 1347, 1355 (Fed. Cir. 2010), *rev'd*, 132 S. Ct. 1289 (2012) (holding, on remand after *Bilski*, that "as applied to the present claims, the 'useful and important clue, an investigative tool,' leads to a clear and compelling conclusion, *viz.*, that the present claims pass muster under § 101" (quoting *Bilski*, 561 U.S. at 604)).

24 *Mayo Collaborative Servs.*, 132 S. Ct. at 1294.

25 Ass'n for Molecular Pathology v. Myriad Genetics, Inc., 133 S. Ct. 2107, 2111 (2013).

26 If a firm cannot fully protect diagnostic methods or algorithms that involve a piece of genetic information, patents on the isolated gene of interest can still prevent others from determining the gene variant and therefore from practicing the method. Myriad Genetics used this strategy to protect its breast cancer diagnostic tests. *Id.* at 2113. This strategy is imperfect; indeed, whole-genome sequencing likely circumvents isolated gene patents, *see* W. Nicholson Price II, *Unblocked Future: Why Gene Patents Won't Hinder Whole Genome Sequencing and Personalized Medicine*, 33 CARDOZO L. REV. 1601 (2012), but blocked many market entrants in *Myriad*'s case.

27 *See* Alice Corp. Pty. v. CLS Bank Int'l, 134 S. Ct. 2347, 2358 (2014).

After *Prometheus*, *Myriad*, and *Alice*, incentives for developing personalized medicine—especially the complex algorithms at the heart of black-box medicine—are much lower than they were before.[28] Perhaps more importantly, the remaining incentives available now push personalized medicine in the wrong direction. Because patents are on stronger ground when they cover inventions that closely link devices or treatments to a new correlation or algorithm, firms are likely to prioritize development of those combination products rather than pursuing broader analyses of large datasets and complex correlations within them. This pushes firms toward maintaining the current model of simple, explicit relationships, rather than developing and exploiting the far larger realm of complex and often implicit relationships.

In addition, firms may increasingly turn away from the patent system and rely instead on trade secrecy law and practices to protect proprietary data and algorithms.[29] Secrecy is problematic for medicine in general, but especially for black-box medicine. Because black-box medicine already involves complex and frequently implicit relationships, as much transparency as possible is needed for validation and oversight. In addition, cumulative innovation based on shared data and algorithms is crucial to advancing the field, but is restrained by pervasive secrecy.

So how can we smooth the path for black-box medicine? The first reaction to inadequate innovation incentives is often to throw more intellectual property protection at the issue to drive it forward, but this Article argues that this approach is insufficiently nuanced here, and that it raises new problems. Developing black-box medicine involves solving multiple interconnected problems: generating and consolidating the necessary data, developing algorithms and models, and validating those models for medical use.[30] Each of those processes requires individual consideration through the lens of innovation policy; while algorithms follow familiar innovation patterns, databases—especially large, broad databases aimed at driving future innovation—are more similar to infrastructure than to inventions, and therefore need to be the subject of substantial public efforts to develop them. Validation of algorithms requires yet another set of incentives, potentially in the form of a

---

28 This is not to say that no incentives exist—first mover advantages, trade secrecy, and whatever patents are available provide some incentives. Nor is to argue that black-box medicine is not being developed at all—some firms are active in the space—but rather that available incentives are smaller than optimal and that black-box medicine is being developed less and more slowly than would be preferable.

29 *See* Barbara J. Evans, *Economic Regulation of Next-Generation Sequencing*, 42 J.L. MED. & ETHICS 51 (2014) [hereinafter Evans, *Economic Regulation of Next-Generation Sequencing*].

30 *See* discussion *infra* Section I.A.

"validation bounty" to encourage third parties to evaluate black-box medicine.

This Article proceeds in four Parts. Part I briefly describes black-box medicine, discusses the hurdles to its development, and discusses the need for policy incentives. Part II addresses the patent incentives available for personalized medicine and the diagnostic tests and algorithms on which it relies. Part III addresses non-patent incentives. Part IV discusses potential solutions and policy interventions. A few brief thoughts conclude.

I. BLACK-BOX MEDICINE, HURDLES, AND THE NEED FOR INCENTIVES

Black-box medicine requires substantial investment to pursue. This Part describes black-box medicine in more detail, then lists the three main types of practical developmental hurdles—data, algorithms, and validation—and concludes by offering the case for providing economic incentives to drive development forward.

A. *Black-Box Medicine*

Medical science has long relied on clinical trials to demonstrate the efficacy of interventions, whether pharmaceutical, surgical, or device-based.[31] Increasingly, however, doctors, patients, scientists, and policy-makers are recognizing that because patients are different from one another, medical interventions should frequently be tailored to the specific characteristics of each individual patient; recent years have seen an increased focus on personalized medicine, which relies on this tailoring to provide "the right drug for the right patient at the right dose and time."[32] In his 2015 State of the Union Address, President Obama announced the Precision Medicine Initiative,[33] a $215 million multiyear initiative to accelerate the development of personalized medicine.[34]

In the early stages of personalized medicine, development has relied on well-validated, explicit, and—as a result—relatively simple links between a particular patient characteristic and the resulting

---

31 *See generally* Rothwell, *supra* note 8.

32 Wolfgang Sadée & Zunyan Dai, *Pharmacogenetics/Genomics and Personalized Medicine*, 14 HUM. MOLECULAR GENETICS (SUPPLEMENT 2) R207, R207 (2005).

33 Address Before a Joint Session of Congress on the State of the Union, *supra* note 3.

34 Collins & Varmus, *supra* note 4; Press Release, White House, Fact Sheet: President Obama's Precision Medicine Initiative (Jan. 30, 2015), https://www.whitehouse.gov/the-press-office/2015/01/30/fact-sheet-president-obama-s-precision-medicine-initiative.

intervention.[35] The poster child of personalized medicine is the use of the drug Herceptin to treat breast cancer. Clinical trials confirm that Herceptin is effective only when the tumor overexpresses the gene for a particular receptor that the drug targets.[36] Women who are considering taking Herceptin can use a test to evaluate their tumor's expression level to make the right treatment decision.[37] Current personalized medicine mostly follows this model: well-understood scientific links between patient characteristics and interventions are validated through clinical trials and then adopted into medical practice.[38]

Unfortunately, this mode of development comes with substantial limitations. Our ability to map biological relationships explicitly is limited, and as a consequence we have more trouble elucidating the complex biological networks that underlay much of human disease.[39] Even more limited is the tool of clinical trials. More complex relationships split people into more precisely defined categories—that is, after all, the end goal of personalized medicine—but that means that fewer people exist in each category, and that those people are harder to find.[40] Such splitting fits poorly with the clinical trial model of aggregating larger numbers of roughly comparable individuals and grouping them into a small number of sets to measure the difference at issue. For validating complex biological relationships, even assuming they can be identified by the underlying science, clinical trials pose major challenges.

Black-box medicine seeks to address these challenges by leveraging the availability of large amounts of health data and the increasing sophistication of machine-learning algorithms.[41] Health data are constantly expanding, including genetic sequences, metabolic screens, and the increasing amount of health information included in electronic health records.[42] Sophisticated algorithms can find patterns in these

---

35 *See, e.g.*, Carney, *supra* note 5.

36 *Id.* at 25–27.

37 *See* John C. Mansour & Roderich E. Schwarz, *Molecular Mechanisms for Individualized Cancer Care*, 207 J. AM. C. SURGEONS 250, 250–58 (2008).

38 *See generally* Price, *Black-Box Medicine*, *supra* note 7, at 427–29.

39 *See, e.g.*, Revathi Rajkumar & Ferhaan Ahmad, *The Genomic Complexity Underlying Pulmonary Arterial Hypertension: From Mendel to Networks*, 189 AM. J. RESPIRATORY & CRITICAL CARE MED. 1152 (2014) (describing complex interactions of genetic and environmental factors); Takanori Watanabe et al., *Disease Prediction Based on Functional Connectomes Using a Scalable and Spatially-Informed Support Vector Machine*, NEUROIMAGE, Aug. 1, 2014, at 183 (using machine-learning techniques to analyze six-dimensional spatial mappings of neuronal connections in the brain to predict schizophrenia).

40 Ultimately, the goal is to base medical decisions on the full picture of each unique individual; in a real sense, then, the eventual number of precisely relevant individuals should converge to one.

41 *See generally* Price, *Black-Box Medicine*, *supra* note 7.

42 *See id.* at 430–31.

data, whether those patterns reflect which patients are likely to benefit most from a limited resource like an inpatient hospital bed,[43] what pattern of 5000 genes predicts how a lung tumor will respond to treatment,[44] or the possibility that a patient with an unremarkable collection of characteristics is actually at very high risk for a rare disease.[45]

While machine-learning algorithms can pull useful relationships out of large datasets, the challenge is that such relationships are frequently opaque.[46] Sometimes, the relationships will be formally opaque—that is, in some machine learning techniques, it is actually impossible to state how the algorithm classifies observations once it has been developed.[47] Other times, the relationships will only be practically opaque—that is, they may be so complicated that they defy explicit understanding, like a relationship between fifty genes, several environmental factors, and a health outcome.[48] In either case, the opacity of the relationships makes them hard to impossible to validate by the traditional methods of scientific understanding and clinical trials.[49]

In sum, black-box medicine is the use of non-transparent computer algorithms to make health-care decisions.[50] It has tremendous

---

43 *See* I. Glenn Cohen et al., *The Legal and Ethical Concerns that Arise from Using Complex Predictive Analytics in Health Care*, 33 HEALTH AFF. 1139, 1140 (2014).

44 Hojin Moon et al., *Ensemble Methods for Classification of Patients for Personalized Medicine with High-Dimensional Data*, 41 ARTIFICIAL INTELLIGENCE MED. 197 (2007).

45 Joseph A. Cruz & David S. Wishart, *Applications of Machine Learning in Cancer Prediction and Prognosis*, 2 CANCER INFORMATICS 59 (2007).

46 *See* Price, *Black-Box Medicine*, *supra* note 7, at 432–34.

47 *See id.* at 433–34.

48 *See id.* at 434.

49 *See id.* at 440–41.

50 More precisely, there are two different types of algorithms involved in black-box medicine. The first is the machine-learning algorithm itself, which identifies patterns; the second is a prediction/recommendation algorithm incorporating insights from the first. These two may be unified or separated. To take a simple example from explicit personalized medicine, consider the case of the cancer drug Herceptin and expression levels of the receptor gene HER2/neu, mentioned above. *See supra* notes 36–37 and accompanying text. Only tumors overexpressing HER2/neu are responsive to Herceptin treatment. *See supra* notes 36–37 and accompanying text. Assuming the relationship was unknown, one could conceive of a search algorithm designed to find the relationship. Such an algorithm could examine a dataset containing expression levels for many genes and tumor responsiveness to Herceptin, calculating correlations between each gene's expression level and responsiveness, returning the strongest correlation—presumably HER2/neu. *See, e.g.*, Erdal Cosgun et al., *High-Dimensional Pharmacogenetic Prediction of a Continuous Trait Using Machine Learning Techniques with Application to Warfarin Dose Prediction in African Americans*, 27 BIOINFORMATICS 1384, 1385–86 (2011) (describing algorithms deployed to find alleles related to warfarin dosing in an African-American cohort). That search would then yield a separate, much simpler treatment-recommendation algorithm: measure the expression level of HER2/neu, and if that level is above a certain threshold, treat with Herceptin; otherwise, pursue a different strategy. *See, e.g.*, *id.* at 1386–87 (discussing performance of resulting models recommending warfarin dosage). In

potential for health care, and raises possibilities for substantial savings both in development (as compared to traditional medical development pathways) and in application (since a closer match between patient and treatment can avoid costly errors and wasted treatment). However, black-box medicine still requires substantial resources to develop.

B. *Hurdles to Development*

Black-box medicine promises the possibility of identifying new treatments and targeting those treatments for substantially lower costs, in time and money, than the current explicit pathway for developing personalized medicine. Nevertheless, black-box medicine still faces substantial practical hurdles in development.[51] This Section addresses three key hurdles. First, black-box medicine requires the assembly of large, high-quality datasets of health information. Second, algorithms themselves must be developed, which requires substantial expertise. Third and finally, predictive relationships must be validated to assure safe and effective use in medical practice.

1. Datasets

The first, and likely most expensive, requirement for black-box medicine is the generation of large, high-quality datasets of health information. A key advantage of black-box medicine is that the expense of data *creation* is not required, because black-box medicine relies principally on sophisticated retrospective analyses. Thus, the extraordinarily expensive process of clinical trials is not necessary. However, the effective use of existing and contemporaneously generated data requires surmounting at least two practical challenges: acquiring

other, more complex or opaque situations, the search and prediction algorithms would be the same; for instance, a neural network trained on a set of data to perform complex classification tasks is both the search algorithm (as it is being trained) and the prediction algorithm (once trained, and in use thereafter). *See, e.g.*, Francesco Ciompi et al., *Automatic Classification of Pulmonary Peri-Fissural Nodules in Computed Tomography Using an Ensemble of 2D Views and a Convolutional Neural Network out-of-the-Box*, MED. IMAGE ANALYSIS, Dec. 2015, at 195 (describing neural networks trained and used in lung cancer screening). For ease of explanation, these two algorithmic functions are elided through the remainder of this Article.

51 Black-box medicine also faces substantial policy and legal questions, including how it will be regulated, how it will be reimbursed, and how privacy and informed consent concerns will be adequately addressed. These policy questions are outside the scope of this article, though they are noted and briefly described in Price, *Black-Box Medicine*, *supra* note 7, at 442–66. To the extent that uncertainty about policy questions decreases the expected benefits to a firm of developing a particular black-box implementation, those concerns reduce innovation incentives as well.

and linking data from different sources, and ensuring the quality of the final dataset.

a. Data Collection

First, firms must gain access to the substantial amounts of data in electronic form. As electronic medical records become more prevalent, collecting data should become practically easier because data must only be collected and translated, not moved from paper records into electronic form. However, paper records will remain significant for the development of black-box medicine, as they are necessary to provide legacy data and information to elucidate longer-term patterns.[52] Other data are new in kind; as broad screening tests, such as whole-genome sequencing or metabolic screens, move into more widespread practice, the data available to be gathered will likewise increase.[53]

Concerns regarding economics and patient-consent are potentially more challenging. Health data are valuable; health care systems know it, and health care providers and patients sometimes do as well.[54] Getting information will often require compensating whoever has gathered the information in the first place—done the testing, compiled the records, or conducted the screening—or the patient whose data is being gathered, or both. In either case, consent from patients will generally be required, either at the initial data collection (which may or may not have already happened) or at the transfer of data.[55]

There is a large exception to the consent requirement: in many situations, the most straightforward way to acquire individual-level patient data is to anonymize that information.[56] However, anonymization runs into two interrelated problems. First,

---

52 *See* Diane Dolezel & Jackie Moczygemba, *Implementing EHRs: An Exploratory Study to Examine Current Practices in Migrating Physician Practice*, PERSP. HEALTH INFO. MGMT., Winter 2015, at 2–3, 13; Roy Schoenberg & Charles Safran, *Internet Based Repository of Medical Records that Retains Patient Confidentiality*, 321 BMJ 1199, 1199 (2000).

53 *See* Antonio Regalado, *EmTech: Illumina Says 228,000 Human Genomes Will Be Sequenced this Year*, MIT TECH. REV. (Sept. 24, 2014), http://www.technologyreview.com/news/531091/emtech-illumina-says-228000-human-genomes-will-be-sequenced-this-year.

54 *See* Barbara J. Evans, *Sustainable Access to Data for Postmarketing Medical Product Safety Surveillance Under the Amended HIPAA Privacy Rule*, 24 HEALTH MATRIX 11 (2014) [hereinafter Evans, *Sustainable Access to Data*]; Marc A. Rodwin, *Patient Data: Property, Privacy & the Public Interest*, 36 AM. J.L. & MED. 586 (2010).

55 Consent is required under the Privacy Rule, 45 C.F.R. § 164.501 (2015), of the Health Insurance Portability and Accountability Act of 1996 (HIPAA), Pub. L. No. 104-191, 110 Stat. 1936 (codified as amended in scattered sections of 18 U.S.C., 26 U.S.C., 29 U.S.C., and 42 U.S.C.).

56 Anonymization is enough to remove the requirement of patient consent for data use under HIPAA. *See* 45 C.F.R. § 164.514(a). For a discussion of the substantial differences between anonymity and privacy, see Jeffrey M. Skopek, *Anonymity, the Production of Goods, and Institutional Design*, 82 FORDHAM L. REV. 1751 (2014).

anonymization is imperfect; even if patient-identifying information is removed from a medical record, the remaining information can be used to re-identify the individual.[57] Second, anonymization complicates the task of assembling data about one patient into integrated records. If data from one source about a particular individual cannot be matched with data from another source about that same individual, substantial information is lost. Similarly, if information from a patient at one point in time cannot be supplemented with later information, risks and benefits are much harder to evaluate and leverage. While technological solutions are feasible,[58] they create an added layer of complexity, and the more robust the mechanism for ensuring that all of an individual's data can in fact be collected in a single record, the greater the chance of re-identification for that individual, both based on the collation mechanism and on the collected health data themselves.[59]

Individual corporations have gathered significant amounts of patients' health data; however, those data are typically jealously guarded and unavailable for others to use in developing medical models.[60] As

---

57 *See, e.g.*, Bradley Malin & Latanya Sweeney, *How (Not) to Protect Genomic Data Privacy in a Distributed Network: Using Trail Re-Identification to Evaluate and Design Anonymity Protection Systems*, 37 J. BIOMEDICAL INFORMATICS 179 (2004); Paul Ohm, *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization*, 57 UCLA L. REV. 1701 (2010); Felix T. Wu, *Defining Privacy and Utility in Data Sets*, 84 U. COLO. L. REV. 1117 (2013); Jane Yakowitz, *Tragedy of the Data Commons*, 25 HARV. J.L. & TECH. 1 (2011). Note that re-identification is a practical and ethical concern more than a legal one, as current regimes typically do not acknowledge the possibility. *See* 45 C.F.R. §§ 164.514(a)–(b)(2). Access to at least some currently anonymous datasets requires an agreement that the requester will not attempt to re-identify the individuals whose data is being shared. *See, e.g.*, *Data Use Restrictions*, CENTERS FOR DISEASE CONTROL & PREVENTION, http://wonder.cdc.gov/DataUse.html (last visited Dec. 22, 2015) ("The *CDC/ATSDR Policy on Releasing and Sharing Data* prohibits linking these data with other data sets or information for the purpose of identifying an individual.").

58 Vanderbilt's eMerge Network follows this model. *See About*, eMERGE NETWORK, https://emerge.mc.vanderbilt.edu/about-emerge (last visited Dec. 22, 2015); *see also* Kristin Madison, *Health Regulators as Data Stewards*, 92 N.C. L. REV. 1605, 1616 (2014) ("Participants in the network agree to submit genetic data to a coordinating center that will then combine the data with the network dataset and submit them to the database of Genotypes and Phenotypes, which makes individual-level genetic data available to researchers.").

59 For instance, if the collation mechanism allows those adding data into a database to access the identity of the data record and the data being added in order to ensure that the record is entirely about the same person, that would be a collation mechanism that could breach anonymity. On the other hand, if the result of a perfectly anonymous collation mechanism is a dataset that contains very large amounts of health data for each referenced individual, that amount of data helps enable re-identification efforts. For an overview of re-identification challenges, see Paul Ohm, *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization*, 57 UCLA L. REV. 1701 (2010).

60 *See, e.g.*, Press Release, 23andMe, 23andMe Announces Collaboration with Pfizer Inc. to Conduct Genetic Research Through 23andMe's Research Platform (Jan. 12, 2015), http://www.prnewswire.com/news-releases/23andme-announces-collaboration-with-pfizer-inc-to-conduct-genetic-research-through-23andmes-research-platform-300018683.html (announcing

described earlier, black-box medicine relies on identifying complex patterns in health datasets. Its strength, both in finding relationships and in verifying that those relationships are real, depends on having large datasets with varied patients. Keeping data separate in corporate data silos, particularly silos created for particular medical or economic purposes,[61] enervates the broader power of black-box medicine.

b. Data Quality

The second major challenge is ensuring the quality of the collected data. Because black-box medicine relies on data-based development and validation, and lacks the potential for specific scientific or clinical validation, data quality is especially important. Errors in data can lead to false pattern recognition, although theoretically, only biased data should create such problems. Data that are afflicted by random error make patterns harder to recognize, however, decreasing the number of relationships that can be discovered and used and requiring larger datasets to observe the same patterns.

As others have noted, problems in data quality can arise in multiple ways.[62] Front-line coders—that is, the doctor entering data in her office or the technician reporting lab results—can introduce simple human error when inputting data into electronic systems.[63] Entering data from paper records creates another opportunity for human error.[64]

In addition to random error, data quality problems can arise due to systemic incentives inherent in the data collection environment. Much biomedical data is collected in and for insurance records. Doctors have incentives to "upcode" treatments to receive higher reimbursement; even in situations that fall short of fraud, ambiguous situations are more likely to be coded as the more expensive alternative.[65] Setting aside

---

a collaboration allowing Pfizer to research lupus using 23andMe's "largest dataset of its kind," including over 800,000 individuals' genotyped samples).

61 Myriad Genetics, for instance, developed a business strategy focused on being the exclusive provider of tests for the BRCA1/2 breast-cancer-predisposition genes. Myriad has a strong interest in keeping other researchers from using its data, and has no incentive to integrate its data into broader health prediction pictures so long as those analyses are conducted or controlled by others. Accordingly, Myriad's extensive data on BRCA1/2 variants and their significance are generally unavailable to others developing predictive algorithms, along with the health data collected on the women who provided the data and, in many cases, their relatives. *See* Dan L. Burk, *Patents as Data Aggregators in Personalized Medicine*, 21 B.U. J. SCI. & TECH. L. 233, at 240–54 (2015).

62 *See, e.g.*, Sharona Hoffman & Andy Podgurski, *The Use and Misuse of Biomedical Data: Is Bigger Really Better?*, 39 AM. J.L. & MED. 497, 515–21 (2013).

63 *Id.*

64 *Id.*

65 Christopher S. Brunt, *CPT Fee Differentials and Visit Upcoding Under Medicare Part B*, 20 HEALTH ECON. 831 (2011). This incentive largely relies on the dominant fee-for-service model, and is likely to be decreased or eliminated in capitated payment plans, where the health

biased incentives, the purpose of insurance data collection—ensuring that appropriate payments are made—means that more emphasis is put on cost-related aspects of treatment, and medically relevant differences may be elided.[66]

Ensuring the greatest possible quality for black-box medicine datasets requires combating these sources of error and bias. To the extent that redundant data can be gathered—for instance, both insurance reimbursement records and electronic health records—those redundant sources can be compared to identify potential coding mistakes. In the process of physically coding information from paper records to electronic datasets, multiple coders can be used and intercoder reliability used as a check on quality.

Overall, assembling the data needed to develop black-box medicine may present the largest cost hurdle to its development. It requires coordination among multiple data sources, checks on quality, and—though not discussed in detail here—compliance with legal and regulatory requirements.[67] These expenses are likely much less than the tremendous costs associated with generating gold-standard clinical trial data, but are nonetheless significant. Good data are the foundation of black-box medicine, and require investment accordingly.

2. Algorithm Development

The second major cost hurdle for black-box medicine is its heart: the development of algorithms to find patterns in the data and then to predict medical outcomes and recommend treatment. Predictive algorithms are increasingly sophisticated—indeed, that sophistication enables the possibility of black-box medicine in the first place—but their development continues to require substantial time, programming

---

care provider is compensated per patient or per episode of care rather than by procedure. James C. Robinson, *Theory and Practice in the Design of Physician Payment Incentives*, 79 MILBANK Q. 149, 158 (2001).

66 *See* Hoffman & Podgurski, *supra* note 62, at 519.

67 Research must comply with HIPAA, especially its Privacy Rule, 45 C.F.R. pts. 160, 162, 164; the Common Rule, 45 C.F.R. pt. 46, which governs human subjects research funded by any of several government agencies; FDA human subjects research rules, if intended to support drug approval, 21 C.F.R. pt. 50; and various other regulatory requirements. *See, e.g.*, KRISTEN ROSATI ET AL., MINI-SENTINEL PRIVACY PANEL, HIPAA AND COMMON RULE COMPLIANCE IN THE MINI-SENTINEL PILOT (2011), http://mini-sentinel.org/work_products/About_Us/HIPAA%20and%20Common%20Rule%20Compliance%20in%20the%20Mini-Sentinel%20Pilot.pdf (describing HIPAA and Common Rule requirements for FDA's Mini-Sentinel pilot project to monitor adverse drug reactions); Evans, *Sustainable Access to Data*, *supra* note 54 (discussing policy issues with FDA's Sentinel project); Price, *Black-Box Medicine*, *supra* note 7, at 454–57 (describing privacy concerns in black-box medicine). A full accounting of these requirements is outside the scope of this Article.

experience, and computational resources. Even simple predictive algorithms, such as movie recommendations for the video-on-demand service Netflix, require major effort to develop and optimize.[68] Relatively small and simple datasets can still run into constraints in terms of computer processing power,[69] and expert programmers are generally required to develop the most appropriate algorithms. In other, more complex fields where predictive algorithms are deployed, such as finance or creditworthiness, even more significant investments are required for algorithmic development.[70]

3. Validation

The third and final cost hurdle for black-box medicine is the requirement for validation to ensure high quality. Black-box medicine lacks the validation of scientific understanding and clinical trials, and thus requires other approaches. Ensuring that black-box algorithms are as well validated as possible based on available data is crucial to improving health-care quality, as well as to promoting provider and patient acceptance.[71]

Algorithms developed exclusively through discerning patterns in complex health data lack the two typical forms of validation used in current models of personalized medicine: scientific understanding and targeted clinical trials.[72] Although some current medical treatments are developed and applied without an understanding of mechanism,[73] in most instances we understand approximately how and why the treatment works. This scientific understanding provides a base-level validation of a treatment option; if we can say that a particular

---

68 *See* Price, *Black-Box Medicine*, *supra* note 7, at 439 (describing the process of Netflix updating its movie recommendation algorithm).

69 *See* ANDREAS TÖSCHER, MICHAEL JAHRER & ROBERT M. BELL, THE BIGCHAOS SOLUTION TO THE NETFLIX GRAND PRIZE 3, 9, 15, 17 (2009), http://www.netflixprize.com/assets/GrandPrize2009_BPC_BigChaos.pdf (noting that some algorithms could only be run a limited number of times due to memory, storage, and processing power limitations).

70 PASQUALE, *supra* note 11.

71 Validation is a challenge from both innovation and regulation points of view, since it impacts both technological development and the assurance of high quality for public use. These two goals are intertwined; the creation of information about new medical technologies is itself a form of innovation. *See* Rebecca S. Eisenberg, *The Role of the FDA in Innovation Policy*, 13 MICH. TELECOMM. & TECH. L. REV. 345 (2007) [hereinafter Eisenberg, *The Role of the FDA in Innovation Policy*].

72 *See* Price, *Black-Box Medicine*, *supra* note 7, at 440–42.

73 For instance, aspirin had an unknown mechanism of action for decades, and lithium's mechanism is still not well understood. *See* Gin S. Malhi et al., *Potential Mechanisms of Action of Lithium in Bipolar Disorder*, 27 CNS DRUGS 135, 136 (2013); J. R. Vane, *Inhibition of Prostaglandin Synthesis as a Mechanism of Action for Aspirin-like Drugs*, 231 NATURE NEW BIOLOGY 232 (1971).

monoclonal antibody treats rheumatoid arthritis because it targets a protein receptor that triggers the inflammatory response, that understanding provides validation that the treatment is at least potentially legitimate and effective.[74] When the relationships are too complex to understand, or are literally opaque, scientific understanding is a priori unavailable as a potential source of validation.

The second typical form of validation, required for FDA approval for most treatment options or diagnostic tests, comes in the form of targeted clinical trials. Different sets of patients receive the specific treatment, and if the treated set of patients showed statistically significant improvements over the control set, the clinical trial validates the treatment. Black-box medicine algorithms are largely not amenable to clinical trials because the underlying relationships are either unknown or too complex to collect sufficiently large patient samples. This is especially true of more holistic algorithms that incorporate multiple relationships into more general predictions and recommendations.

In the development and validation of black-box medicine, large unbiased samples can partially substitute for the methodological strength of randomized clinical trials, but some form of validation is still required to ensure the strength, accuracy, and quality of the resulting algorithm. Because the traditional forms of validation are unavailable, black-box medicine should rely instead on efforts by external parties to computationally support the original algorithm, both by evaluating the development parameters (i.e., how the algorithm was developed) and by trying to independently reach similar results through parallel computational methods—ideally on parallel data.[75] This validation effort, which is closely tied to regulation for quality, is nonetheless a form of costly innovation, and a hurdle that black-box medicine will need to overcome.

Overall, the costs and hurdles associated with developing black-box medicine are significant and require major investment from relevant stakeholders. This does not in itself justify policy intervention to increase incentives; in many situations, the market is expected to reward the need for large investment, so that no particular policy action need be taken. The next Section briefly makes the case that additional incentives are required for black-box medicine's development.

---

74 Nancy J. Olsen & C. Michael Stein, *New Drugs for Rheumatoid Arthritis*, 350 NEW ENG. J. MED. 2167, 2170–75 (2004).

75 *See* Price, *Black-Box Medicine*, *supra* note 7, at Section II.C.

C. *The Need for Incentives*

The basic justification for intellectual property is a well-told story.[76] Society derives tremendous benefits from innovation, but absent intellectual property, ideas are frequently expensive to produce but hard to protect.[77] In addition to the initial discovery, the process of taking an idea through the development into a commercial product can be costly and is frequently subject to free-riding.[78] Since firms cannot capture much of the value of their investments in innovation, they invest at a socially suboptimal level.[79] Intellectual property allows firms to capture some of that surplus, increasing the incentives for invention by allowing firms to exclude others from the invention.[80] Black-box medicine follows this pattern closely: databases, algorithms, and the knowledge that algorithms are reliable are all information goods, which are difficult to keep exclusive once known. Accordingly, intellectual property—or a substitute incentive set—is likely necessary for its socially optimal development.[81]

The patent system fills this role by guaranteeing significant protection in exchange for disclosure of the technology, thus increasing the type of protection available and enabling more cumulative innovation.[82] Patents provide an alternative to either not developing an appropriable innovation or keeping it secret. Secrecy prevents appropriation and is bolstered by the mostly state-law doctrine of trade

---

76 *See, e.g.*, Mark A. Lemley, *Ex Ante Versus Ex Post Justifications for Intellectual Property*, 71 U. CHI. L. REV. 129, 129 (2004) [hereinafter Lemley, *Ex Ante Versus Ex Post Justifications for Intellectual Property*]; Kevin Outterson, *Pharmaceutical Arbitrage: Balancing Access and Innovation in International Prescription Drug Markets*, 5 YALE J. HEALTH POL'Y L. & ETHICS 193, 195–98 (2005); Benjamin N. Roin, *Unpatentable Drugs and the Standards of Patentability*, 87 TEX. L. REV. 503, 507–09 (2009) [hereinafter Roin, *Unpatentable Drugs and the Standards of Patentability*].

77 *See* Lemley, *Ex Ante Versus Ex Post Justifications for Intellectual Property*, *supra* note 76, at 129.

78 *Id.*

79 *Id.*

80 *Id.*

81 *See* Peter S. Menell, *Tailoring Legal Protection for Computer Software*, 39 STAN. L. REV. 1329 (1987) (describing computer software as a public good with positive network externalities and suggesting government intellectual property interventions to encourage its development).

82 Patents also hamper cumulative innovation, if the patent on the original invention blocks the second innovator from developing her innovation; the extent to which this occurs is something of an open empirical question. *See generally* Suzanne Scotchmer, *Standing on the Shoulders of Giants: Cumulative Research and the Patent Law*, J. ECON. PERSP., Winter 1991, at 29; *see also* Arti K. Rai, *Fostering Cumulative Innovation in the Biopharmaceutical Industry: The Role of Patents and Antitrust*, 16 BERKELEY TECH. L.J. 813, 838–44 (2001) (describing how patent doctrine can facilitate cumulative innovation). However, effective trade secrecy can very effectively prevent cumulative innovation because the mechanics of the initial innovation never become known.

secrecy,[83] but works poorly for innovations which can be reverse-engineered or which are unavoidably public.[84] In addition, trade secrecy limits cumulative innovation, where different innovators build on the inventions and innovations of other firms.[85]

Overall, black-box medicine is a promising branch of personalized medicine that offers significant advances, but also requires significant investment in nonexcludable goods. Due to its nature as a public information good, firms are likely to invest in black-box medicine below socially desirable levels. Accordingly, innovation incentives should be provided at the policy level. The next Part turns to existing patent incentives offered by the intellectual property system, and failures in those incentives to drive the development of black-box medicine.[86]

II. PATENT INCENTIVES

Patents are a key policy tool to drive technological innovation, and are particularly important in the biomedical fields, playing a crucial role in the development of new drugs and biologics.[87] Patents have also been the subject of significant dispute in those fields.[88] In general, therefore,

---

83 For a general overview of trade secrecy law, see Robert G. Bone, *A New Look at Trade Secret Law: Doctrine in Search of Justification*, 86 CAL. L. REV. 241, 247–51 (1998).

84 *See* RESTATEMENT (THIRD) OF UNFAIR COMPETITION § 43 (AM. LAW INST. 1995) (listing reverse engineering as a proper means of acquiring a trade secret). Trade secrecy in the context of black-box medicine will be discussed below in Section III.C.2.

85 *See* Bone, *supra* note 83, at 266–67.

86 One important caveat is that this description of incentives applies most cleanly to well-defined, relatively stable algorithms—that is, algorithms that are developed once and then used for some time. Black-box medicine offers the possibility of more plastic algorithms, however, that are constantly updated as new information is received. A full analysis of the incentive implications of such flexible second-generation algorithms is beyond the scope of this Article.

87 *See, e.g.*, Rebecca S. Eisenberg, *The Problem of New Uses*, 5 YALE J. HEALTH POL'Y L. & ETHICS 717 (2005); Roin, *Unpatentable Drugs and the Standards of Patentability*, *supra* note 76. Many commentators have also criticized this view and the dominance of pharmaceutical patents. *See, e.g.*, Ellen 't Hoen, *TRIPS, Pharmaceutical Patents, and Access to Essential Medicines: A Long Way from Seattle to Doha*, 3 CHI. J. INT'L L. 27 (2002); Tim Hubbard & Jamie Love, *Medicines Without Barriers*, NEW SCIENTIST, June 14, 2003, at 29.

88 Various policy arguments around patents have included the use of patents to extend drug monopolies for longer terms than contemplated in the patent term, *see, e.g.*, C. Scott Hemphill & Bhaven N. Sampat, *Evergreening, Patent Challenges, and Effective Market Life in Pharmaceuticals*, 31 J. HEALTH ECON. 327 (2012); alleged antitrust violations when brand-name drug companies and generic companies agree to delay generic market entry in patent litigation, *see, e.g.*, Lisa Allen, Note, *Reviewing the Legality of Pharmaceutical Reverse Payment Settlements: The FTC Doesn't Get It Right*, 8 GEO. J.L. PUB. POL'Y 245 (2010); Daniel A. Crane, *Per Se Illegality for Reverse Payment Patent Settlements*, 61 ALA. L. REV. 575 (2010); Ronald W. Davis, *Reverse Payment Patent Settlements: A View into the Abyss, and a Modest Proposal*, 21 ANTITRUST 26 (2006); the Supreme Court's take on the issue in *Federal Trade Commission v. Actavis, Inc.*, 133 S. Ct. 2223 (2013) (holding that such settlements should be scrutinized under antitrust's rule of reason); and international patent protection hindering patients' access to

patents might appear to provide at least some incentives for the development of algorithms central to black-box medicine, as well as potentially data and validation. However, recent changes to patent subject-matter eligibility law have severely limited those incentives, leaving an incentive landscape that drives personalized medicine away from black-box medicine.[89] In addition, persistent concerns about black-box medicine's ability to meet the written description and enablement provisions of section 112 of the Patent Act make patents more challenging to obtain even if subject-matter eligibility concerns were to be overcome.[90]

A. *Subject Matter Eligibility*

The first question arising for black-box medicine is whether its key innovations can be patented at all. Under 35 U.S.C. § 101, patents can be granted for machines, manufactures, processes, or compositions of matter.[91] Although these broad categories embrace "anything under the sun that is made by man,"[92] they are not infinitely broad, and in particular, they fail to address two of the three key technological hurdles related to black-box medicine. Facts and data do not fall within one of the four categories of patentable subject matter, and thus the collected data enabling black-box medicine are unpatentable. Similarly, the result of validation—that is, whether an algorithm works or not, and how well it might work—is similarly outside the scope of patentable subject matter. This leaves only the algorithms that actually drive black-box medicine as potential subjects of patent protection. Here, however, judicial exceptions to patentability come into play. The Supreme Court has articulated three exceptions to patentable subject matter: abstract ideas, natural phenomena, and laws of nature.[93] In a string of cases, and especially in the 2012 case of *Mayo v. Prometheus*, the Court has made it

lifesaving drugs in developing nations, *see, e.g.*, Amir Attaran & Lee Gillespie-White, *Do Patents for Antiretroviral Drugs Constrain Access to AIDS Treatment in Africa?*, 286 J. AM. MED. ASS'N 1886 (2001); Bryan C. Mercurio, *TRIPS, Patents, and Access to Life-Saving Drugs in the Developing World*, 8 MARQ. INTELL. PROP. L. REV. 211 (2004); Sigrid Sterckx, *Patents and Access to Drugs in Developing Countries: An Ethical Analysis*, 4 DEVELOPING WORLD BIOETHICS 58 (2004). Patents for surgical techniques also raised a furor when they were introduced, but have since been statutorily limited. *See* Robert M. Portman, *Legislative Restriction on Medical and Surgical Procedure Patents Removes Impediment to Medical Progress*, 4 U. BALT. INTELL. PROP. L.J. 91 (1996); *see also* 35 U.S.C. § 287(c) (2012).

89 *See* discussion *infra* Section II.A.

90 *See* discussion *infra* Section II.B.

91 35 U.S.C. § 101 (2012).

92 Diamond v. Chakrabarty, 447 U.S. 303, 309 (1980).

93 Diamond v. Diehr, 450 U.S. 175, 185 (1981).

clear that personalized medicine algorithms will be brought within the ambit of patentability only with great difficulty.[94]

1. Algorithms Before *Mayo v. Prometheus*

Prior to the Supreme Court's *Prometheus* decisions in 2012, patenting diagnostic models was much easier.[95] Although longstanding judicial exceptions prohibited patenting of laws of nature, natural phenomena, or abstract ideas,[96] firms could patent broad methods of treatment where the novel contribution was the newly discovered underlying biological relationship. Algorithms standing alone may be abstract ideas, and biological correlations alone may be laws of nature, but putting those into a broader method claim was relatively easy.[97] Thus, essentially all uses of the algorithm could be protected.

Before *Prometheus*, the Federal Circuit had held generally that a broad diagnostic method was patentable so long as it was either linked to a machine, or resulted in a transformation of matter (the "machine or transformation" test).[98] This test was disapproved in *Bilski* but remained an "important clue" to patentability[99] and generally supported the patentability of diagnostic tests until *Prometheus*. Thus, patent incentives were typically available for diagnostic algorithms.[100]

All this is not to say that patents provided ideal incentives for the algorithm development at the heart of black-box medicine. Complex and especially implicit algorithms are more difficult to describe sufficiently than other inventions, making it harder to satisfy section

---

94 Mayo Collaborative Servs. v. Prometheus Labs., Inc., 132 S. Ct. 1289 (2012). For a detailed description of this line of cases and its application to diagnostic tests, see Rebecca S. Eisenberg, *Diagnostics Need Not Apply*, 21 B.U. J. SCI. & TECH. L. 256 (2015).

95 *See* Eisenberg, *supra* note 94.

96 *Diehr*, 450 U.S. at 185.

97 Extensive case history and subsequent scholarship treats the patentability of algorithms. *See, e.g.*, *id.*; Parker v. Flook, 437 U.S. 584 (1978); Gottschalk v. Benson, 409 U.S. 63 (1972); Donald S. Chisum, *The Patentability of Algorithms*, 47 U. PITT. L. REV. 959 (1986) (describing the doctrine surrounding algorithm patentability). *Flook* and *Diehr* are "difficult to reconcile," Rebecca S. Eisenberg, *Prometheus Rebound: Diagnostics, Nature, and Mathematical Algorithms*, 122 YALE L.J. ONLINE 341, 343 (2013), but this Article focuses on case law that is more recent and directly on-point.

98 *See* Classen Immunotherapies, Inc. v. Biogen Idec, 659 F.3d 1057 (Fed. Cir. 2011); Ass'n for Molecular Pathology v. U.S. Patent & Trademark Office, 653 F.3d 1329 (Fed. Cir. 2011); Prometheus Labs., Inc. v. Mayo Collaborative Servs., 628 F.3d 1347 (Fed. Cir. 2010); Metabolite Labs., Inc. v. Lab. Corp. of Am. Holdings, 370 F.3d 1354 (Fed. Cir. 2004).

99 Bilski v. Kappos, 561 U.S. 593, 603–04 (2010).

100 Of course, the patentable subject matter inquiry is only part of the patentability inquiry. The claimed invention must also be useful, novel, non-obvious, and enabled. *See* 35 U.S.C. §§ 101–103, 112 (2012).

112's written description requirement.[101] In addition, complex algorithm patents are hard to enforce, especially when those algorithms are embedded in medical practice, due to difficulties detecting when the patented algorithm is being used.[102] However, the patent incentives for diagnostic algorithms, while imperfect, were at least still available prior to 2012.

2. *Mayo v. Prometheus*

The Supreme Court's decision in *Prometheus* dramatically changed this situation by holding diagnostic methods essentially unpatentable.[103] *Prometheus* involved two patents of the type described above related to the use of thiopurine drugs to treat autoimmune diseases.[104] Claim 1 of patent 6,355,623, which both the Federal Circuit and the Supreme Court took as exemplary,[105] claims

> [a] method of optimizing therapeutic efficacy for treatment . . . comprising: (a) administering a drug . . . and (b) determining the level of [the metabolite] . . . wherein [a metabolite level below a certain threshold] indicates a need to increase the amount of said drug . . . and wherein [a metabolite level above a different threshold] indicates a need to decrease the amount of said drug subsequently administered to [the] subject.[106]

Prometheus Laboratories (Prometheus), the exclusive licensee of the patents, sells diagnostic kits that embody the patented process.[107] Mayo Clinic Rochester and Mayo Collaborative Services (collectively, Mayo) bought those kits and used them until 2004, when Mayo decided to start making, using, and selling its own kits, with slightly different metabolite level limits.[108] Prometheus brought an infringement action in district court, which held the patents infringed but invalid as claiming a natural law.[109] The Federal Circuit reversed on the grounds that the patents' "administering" and "determining" steps satisfied the "machine or

---

101 *See* discussion *infra* Section II.B.

102 *See* Amy Kapczynski & Talha Syed, Essay, *The Continuum of Excludability and the Limits of Patents*, 122 YALE L.J. 1900 (2013) (noting the difficulty in enforcing patents on information about which treatments do not work).

103 *See* Mayo Collaborative Servs. v. Prometheus Labs., Inc., 132 S. Ct. 1289, 1294, 1297–98 (2012).

104 *Id.* at 1294; U.S. Patent No. 6,355,623 (filed Apr. 8, 1999); U.S. Patent No. 6,680,302 (filed Dec. 27, 2001).

105 *See Prometheus*, 132 S. Ct. at 1295.

106 U.S. Patent No. 6,355,623 (filed Apr. 8, 1999).

107 *Prometheus*, 132 S. Ct. at 1295.

108 *Id.* at 1295–96.

109 *Id.* at 1296.

transformation" test, and therefore, held that the patents did not encompass laws of nature.[110]

On certiorari, the Supreme Court held the invention was not patentable subject matter under section 101 of the Patent Act.[111] Rather than a "genuine application" of an unpatentable natural law, the court said that the patent was invalid because it merely provides the natural law and "tells doctors to engage in well-understood, routine, conventional activity"—namely, measuring metabolite levels and then using Prometheus's new information to inform treatment decisions.[112] Accordingly, the patent was invalid.[113] But this describes the majority of medical diagnostics and black-box medicine.

Under the decision's strikingly broad general analysis, many, if not most, biological diagnostic tests can be characterized as only involving steps that measure levels of biological molecules and then relating that measurement to an underlying natural connection to provide information about the patient's biological characteristics, including genes and their expression levels.[114] Under *Prometheus*, "routine, obvious" pre- or post-solution activity cannot make a claim patentable if it is primarily directed to a law of nature; thus, combining diagnostic methods with standard practice procedures will typically not aid patentability.[115] Notably, though the correlation in *Prometheus* was quite simple, nothing in the opinion limits it to simple relationships, and the Court explicitly eschewed choosing among different laws of nature.[116]

---

110 Prometheus Labs., Inc. v. Mayo Collaborative Servs., 628 F.3d 1347, 1355 (Fed. Cir. 2010). The full procedural history is somewhat more complex. The Federal Circuit held in 2009 that the patents claimed patentable subject matter under its then-dispositive "machine or transformation test." Prometheus Labs. Inc. v. Mayo Collaborative Servs., 581 F.3d 1336 (Fed. Cir. 2009). The Supreme Court granted certiorari, vacated, and remanded with instructions to reconsider in light of its holding in *Bilski v. Kappos*, 561 U.S. 593 (2010), that "the machine or transformation test" was not dispositive, but was merely an important and useful clue to the patentable subject matter inquiry. Mayo Collaborative Servs. v. Prometheus Labs., Inc., 561 U.S. 1040 (2010). On remand, the Federal Circuit found the satisfaction of the test sufficient as a clue to patentability and again held the patents as claiming patentable subject matter. *Prometheus*, 628 F.3d at 1355, *rev'd*, 132 S. Ct. at 1296.

111 *Prometheus*, 132 S. Ct. at 1298 (citing 35 U.S.C. § 101 (2012)).

112 *Id.*

113 *Id.*

114 *See, e.g.*, U. I. Schwarz, *Clinical Relevance of Genetic Polymorphisms in the Human CYP2C9 Gene*, 33 EUR. J. CLINICAL INVESTIGATION (SUPPLEMENT S2) 23, 23–30 (2003).

115 *Prometheus*, 132 S. Ct. at 1298.

116 *Id.* at 1303 ("[O]ur cases have not distinguished among different laws of nature according to whether or not the principles they embody are sufficiently narrow. . . . [T]he cases have endorsed a bright-line prohibition against patenting laws of nature, mathematical formulas and the like . . . .").

Thus, the complexity of relationships in black-box medicine is unlikely to make them patentable under section 101.[117]

The U.S. Patent and Trademark Office has issued guidance to its examiners espousing a broad interpretation of *Prometheus*,[118] and both district courts and the Federal Circuit have invalidated diagnostic test patents based on *Prometheus*.[119] Since most, if not all, diagnostic tests center on identifying new laws of nature and inserting them into the normal flow of clinical practice, *Prometheus* strikes directly at the

---

117 One potential alternate analysis would ask whether the interventions suggested by black-box medicine would themselves be non-routine, which could help a tailored medical treatment patent pass the *Prometheus* test. But this analysis would apply only to some methods. For instance, a black-box algorithm's suggestion to address the risk of a stroke by prescribing an antidepressant might be viewed as a non-routine intervention, but an algorithm that merely identified a buried risk of stroke and suggested normal treatment would likely be viewed as engaging in routine activity. The uncertainty of this analysis could reduce patent incentives even if some broader methods might still be patentable.

118 *See* Memorandum from Andrew H. Hirshfeld, Deputy Comm'r for Patent Examination Policy to the Patent Examining Corps, 2014 Procedure for Subject Matter Eligibility Analysis of Claims Reciting or Involving Laws of Nature/Natural Principles, Natural Phenomena, and/or Natural Products (Mar. 4, 2014) [hereinafter 2014 USPTO Subject Matter Eligibility Guidance], http://www.uspto.gov/patents/law/exam/myriad-mayo_guidance.pdf. This guidance emphasizes that "all claims . . . reciting or involving laws of nature/natural principles, natural phenomena, and/or natural products" must be analyzed using a three-part method:

> [(i)] Is the claimed invention directed to one of the four statutory patent-eligible subject matter categories: process, machine, manufacture, or composition of matter? . . . [(ii)] Does the claim recite or involve one or more judicial exceptions? . . . Judicial exceptions include abstract ideas, laws of nature/natural principles, natural phenomena, and natural products. . . . [(iii)] Does the claim as a whole recite something *significantly different* than the judicial exception(s).

Id. at 2–3.

119 In *PerkinElmer v. Intema*, the Federal Circuit held invalid a claim over a test that established the risk of fetal Down syndrome. PerkinElmer, Inc. v. Intema Ltd., 496 F. App'x 65, 71 (Fed. Cir. 2012). The claimed methods compared marker measurements with each other to predict the risk of Down syndrome. *Id.* at 69. The claim was analogous to that in *Prometheus*, and thus invalid, because it merely claimed a co-occurrence between biological molecules and a natural statistical relationship. *Id.* at 71. Similarly, in *SmartGene v. Advanced Biological Laboratories*, a nonprecedential opinion, the Federal Circuit held unpatentable a system paradigmatic of black-box medicine-type diagnostics, though it relied on expert rules rather than implicit relationships. SmartGene, Inc. v. Advanced Biological Labs., SA, 555 F. App'x 950 (Fed. Cir. 2014). In the patent at issue, the representative claim 1 recited the steps of "(a) providing patient information to a computing device comprising" of three different knowledge bases: "therapeutic treatment regimens," expert rules and advisory information useful for the treatment of a particular disease or medical condition; "(b) generating . . . a ranked listing of available therapeutic treatment regimens . . . ; and (c) generating . . . advisory information." *Id.* at 951–52. The Federal Circuit held that the patent covered abstract ideas, relying both on prior Federal Circuit precedent and on *Prometheus*. *See id.* at 954–56. For further discussion, see Timo Minssen & David Nilsson, *The US Supreme Court in* Mayo v. Prometheus*—Taking the Fire from or to Biotechnology and Personalized Medicine?*, 2 QUEEN MARY J. INTELL. PROP. 376, 383 (2012).

patentability of diagnostics and the personalized medicine of which they are an integral part.[120]

Other recent Supreme Court precedent has bolstered the conclusion that diagnostic and black-box medicine innovations are often unpatentable. In 2013, the Supreme Court held in *Association for Molecular Pathology v. Myriad Genetics* that isolated DNA sequences and other isolated natural phenomena are unpatentable.[121] While the decision did not directly address medical algorithms, patents on the raw materials of medical analyses (genetic sequences, metabolites, RNA, and the like) could have potentially provided a complement to now-unavailable patents on the analyses—but no longer. The *Myriad* patents also claimed genetic diagnostic methods, which the Federal Circuit had already held to be unpatentable subject matter under *Prometheus*.[122] Finally, in 2014, *Alice Corp. v. CLS Bank International* strengthened *Prometheus* by clarifying that abstract inventions, such as algorithms, do not become patentable merely because they are implemented on a computer.[123] Perhaps more importantly, *Alice* vigorously reaffirmed the principles and broad reach of *Prometheus*.[124] *Prometheus*, buttressed by *Myriad* and *Alice*, is now likely to drive firms toward modest, explicit improvements, and away from black-box medicine.

3. The Impact of *Prometheus* on Personalized Medicine

Prometheus has major real-world effects on the industry and black-box medicine. In addition to general negative reactions—such as decreased venture capital investments in the diagnostic industry and pessimistic outlooks[125]—firms may shift product focuses to those which can still be successfully protected by patents. In addition, trade secrecy, with its problems for oversight and cumulative innovation, becomes more attractive by comparison.[126]

---

120 *See, e.g.*, Minssen & Nilsson, *supra* note 119, at 384 ("[C]laims that are broadly directed to what may be considered to be a typical method exploited in personalized medicine will probably be held to be unpatentable under the Prometheus principles."). This is not to argue that the previous patent system created ideal incentives for black-box medicine, a point discussed further below. However, under prior law, at least some patent protection was available.

121 Ass'n for Molecular Pathology v. Myriad Genetics, Inc., 133 S. Ct. 2107, 2111 (2013).

122 Ass'n for Molecular Pathology v. U.S. Patent & Trademark Office, 689 F.3d 1303, 1335 (Fed. Cir. 2012).

123 *See* Alice Corp. Pty. v. CLS Bank Int'l, 134 S. Ct. 2347, 2358 (2014).

124 *See id.* Even if *Prometheus* were tightly limited, it would still cast doubt on the patentability of the sort of diagnostic tests it directly addressed.

125 *See* Heidi Ledford, *Software Patents Await Legal Fate*, 507 NATURE 410, 410 (2014).

126 *See* discussion *infra* Section III.A.

Because *Prometheus* makes natural laws and routine applications of those laws unpatentable, firms may seek to strongly link newly discovered natural laws to machines or to specific treatments. Indeed, firms have done just that by increasing their emphasis on combination products that pair a diagnostic with a device or drug, or both.[127] Since such pairings involve substantially more than just stating a natural law, they are likely still patent-eligible under *Prometheus*.[128] Combination products tend to focus on simple, explicit links, and are tested and brought through an FDA approval process that focuses on validating those links in clinical trials.[129] Thus, while this change in focus by firms may be entirely rational, it means that the contours of intellectual property rights are pushing to keep the industry focused on explicit personalized medicine, rather than devoting energy to the broader algorithms, models, and datasets necessary to bring about the benefits of black-box medicine.

The major problem with moving to a combination product model is that it keeps personalized medicine firmly locked into the current regime of incremental steps. This is not to disparage the potential benefit of combination devices or explicit personalized medicine in general. However, to the extent that firms attempt to maintain patentability by focusing on simple, explicit links associated with devices, they leave untapped the larger datasets and more complex algorithms needed for black-box medicine and its leverage of complex biological relationships. Reliance on combination products also limits the development of medical algorithms largely to the pharmaceutical and medical device industries, which have the capability to market such combination products. Incentives for other entities—health-care payers, informatics companies, or other parties that do not sell drugs—are lower, and those firms may be less likely to innovate as a result.

In sum, though *Prometheus* and its kin may or may not be justified on substantive patent law grounds—a debate into which this Article does not wade—those cases seriously decrease the patent incentives available in the United States for personalized medicine in general, and for black-box medicine in particular.[130]

---

127 *See* Aaron S. Kesselheim & Jason Karlawish, *Biomarkers Unbound—The Supreme Court's Ruling on Diagnostic-Test Patents*, 366 NEW ENG. J. MED. 2338, 2340 (2012).

128 *See* 2014 USPTO Subject Matter Eligibility Guidance, *supra* note 118.

129 U.S. FOOD & DRUG ADMIN., U.S. DEP'T OF HEALTH & HUMAN SERVS., IN VITRO COMPANION DIAGNOSTIC DEVICES: GUIDANCE FOR INDUSTRY AND FOOD AND DRUG ADMINISTRATION STAFF (2014), http://www.fda.gov/downloads/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm262327.pdf. In addition, even if the algorithm is not patentable as part of the combination product, patents on the other part of the combination—the drug or device—may obviate the need for an algorithm patent.

130 This Article considers only domestic protection and incentives; international analyses are beyond its scope. The situation in Europe appears to differ substantially; methods like those at

B. *Section 112*

Section 101's patentable subject matter bar is not the only barrier to patent availability for black-box medicine. A more prosaic hurdle comes from the enablement, written description, and definiteness bars of section 112.[131] These three requirements create substantial hurdles for patenting black-box medicine, and though they do not eliminate the possibility of patenting those inventions which survive the section 101 analysis described above, they further limit the availability of patent incentives for black-box medicine.

Section 112 states that a patent

> shall contain a written description of the invention, and of the manner and process of making and using it, in such full, clear, concise, and exact terms as to enable any person skilled in the art . . . to make and use the same . . . . The specification shall conclude with one or more claims particularly pointing out and distinctly claiming the subject matter which the inventor . . . regards as the invention.[132]

---

issue in *Prometheus* would likely be patentable subject matter under Article 52 of the European Patent Convention (EPC), because they involve in vitro diagnostic tests performed on human subjects. *See* Minssen & Nilsson, *supra* note 119, at 385–86; Paul Cole, *Guest Post:* Prometheus v Mayo—*A European View*, PATENTLY-O (Apr. 3, 2012), http://patentlyo.com/patent/2012/04/guest-post-prometheus-v-mayo-a-european-view.html. In fact, patents on the same methods as in *Prometheus* were granted by the European Patent Office (EPO). *Id.* Those patents were not the subject of opposition proceedings in the EPO, which therefore, did not rule on their patentability (nor, to the author's knowledge, has the EPO ruled on the patentability of precisely analogous claims). However, Article 52(2) of the EPC states that "discoveries, scientific theories and mathematical methods . . . [and] schemes, rules and methods for performing mental acts" are not patentable inventions. Convention on the Grant of European Patents art. 52(2), Oct. 5, 1973, 13 I.L.M. 276 [hereinafter EPC art. 52(2)]. Thus, it is unclear whether purely algorithmic innovations would be patentable. Even if such algorithms are patentable in Europe, their inability to receive patents in the United States increases incentives for firms to keep them secret, or pursue other innovations.

131 *See* 35 U.S.C. § 112 (2012). This argument has been addressed briefly in W. Nicholson Price II, *Describing Black-Box Medicine*, 21 B.U. J. SCI. & TECH. L. 347 (2015) [hereinafter Price, *Describing Black-Box Medicine*]. To receive a patent, a black-box medicine implementation must also satisfy the novelty requirement of § 102, 35 U.S.C. § 102, and the nonobviousness requirement of § 103. 35 U.S.C. § 103. However, those requirements are not substantially different in kind for black-box medicine than for other types of inventions. Potentially, the opacity of an innovation might make it more difficult to determine the exact contours of novelty, and to determine whether the invention was in fact disclosed in the prior art, but that difficulty arises from the § 112 challenges discussed in this section, and therefore will not be treated separately.

132 35 U.S.C. § 112(a)–(b). Section 112 also requires that the patent applicant disclose the best mode of practicing the invention, but under the Leahy-Smith America Invents Act, failing to meet this requirement is no longer grounds for invalidating a patent. *See* Leahy-Smith America Invents Act, Pub. L. No. 112-29, sec. 15, § 282(b)(3)(A), 125 Stat. 284 (2011). Accordingly, this toothless requirement will not be addressed here.

This section has been interpreted to comprise three requirements: first, the invention must be adequately described (written description); second, the patent specification must enable others to practice the invention (enablement); and third, the invention must be described in such specific and definite terms as to clearly lay out the bounds of what is claimed (definiteness).[133] The first two requirements are typically grouped as a disclosure requirement, which serves the dual purposes of informing others about the invention and of limiting the scope of what is claimed.[134] However, "[a]lthough there is often significant overlap between [them], they are nonetheless independent of each other."[135] The next three subsections consider each requirement's challenges for patenting black-box medicine.

1. Enablement

The enablement requirement traditionally provided the bulk of section 112's impact.[136] Under this requirement, the patent specification must enable a person having ordinary skill in the art (the PHOSITA) to practice the invention without "undue experimentation."[137] For inventions requiring biological materials that are not reproducible without undue experimentation, the requirement can be met by making the materials available by depositing them in a public repository.[138]

Satisfying the enablement requirement for black-box medicine algorithms is certainly feasible, but creates significant limits on the strength of those patents. Enablement serves both to require

---

133 *See* Timothy R. Holbrook, *Possession in Patent Law*, 59 SMU L. REV. 123, 123–30 (2006) (summarizing section 112's requirements).

134 *See, e.g.*, MagSil Corp. v. Hitachi Global Storage Techs., Inc., 687 F.3d 1377, 1380–81 (Fed. Cir. 2012) (noting that the enablement requirement "serves the dual function in the patent system of ensuring adequate disclosure of the claimed invention and of preventing claims broader than the disclosed invention"). *See generally* Jason Rantanen, Essay, *Patent Law's Disclosure Requirement*, 45 LOY. U. CHI. L.J. 369, 378–81 (2013) (arguing that the two purposes are actually closely interrelated).

135 Univ. of Rochester v. G.D. Searle & Co., 358 F.3d 916, 921 (Fed. Cir. 2004).

136 *See* Margaret Sampson, Comment, *The Evolution of the Enablement and Written Description Requirements Under 35 U.S.C. § 112 in the Area of Biotechnology*, 15 BERKELEY TECH. L.J. 1233 (2000).

137 *In re* Wands, 858 F.2d 731, 737 (Fed. Cir. 1988). In *Wands*, the court described several factors used to determine whether necessary experimentation is undue,

> includ[ing] (1) the quantity of experimentation necessary, (2) the amount of direction or guidance presented, (3) the presence or absence of working examples, (4) the nature of the invention, (5) the state of the prior art, (6) the relative skill of those in the art, (7) the predictability or unpredictability of the art, and (8) the breadth of the claims.

Id.

138 *Id.* at 735.

disclosure—its ostensible raison d'etre—and to limit the scope of claims.[139] The first of these functions is relatively easy for black-box medicine, at least in principle. For algorithms that are formally transparent and opaque-through-complexity, fully stating the algorithm in the patent will likely enable a PHOSITA to practice that algorithm. For algorithms that are formally opaque—that is, the product of opaque machine-learning algorithms—enablement for that exact algorithm may be achieved by depositing the data and machine-learning algorithm in a publicly available database in the same manner as biological inventions.[140] Nevertheless, this may be an unpalatable solution for innovators seeking to keep their datasets proprietary and to guard their machine-learning algorithms, and also raises substantial privacy concerns.[141]

The greater sting of the enablement requirement comes from the rule that the scope of the claims must be commensurate with the scope of the enablement; that is, an inventor cannot claim more than she has enabled in the patent.[142] To claim a class, an inventor must be able to generalize some aspects of the invention, and to justify that generalization by reference to what members of the class have in common.[143] The opacity of black-box medicine algorithms, whether formal or practical, means that inventors will typically be unable to generalize to broaden claims beyond those algorithms that are precisely disclosed in the specification. This may be a justifiable result, but others will much more easily be able to invent around any covered algorithms, leading to weaker patents—and, consequently, weaker patent-provided incentives for initial innovation.[144]

---

139 ROBERT P. MERGES & JOHN FITZGERALD DUFFY, PATENT LAW AND POLICY: CASES AND MATERIALS 263–64 (6th ed. 2013).

140 *Cf. id.*

141 *See* Price, *Black-Box Medicine*, *supra* note 7, at 454–57. While the HIPAA places substantial restrictions on the use and disclosure of personal health information, it does not apply to anonymous information, so anonymous deposition would resolve HIPAA concerns. *See* 45 C.F.R. § 164.514(a) (2015).

142 *In re* Moore, 439 F.2d 1232, 1236 (C.C.P.A. 1971) ("The relevant inquiry may be summed up as being whether the scope of enablement provided to one of ordinary skill in the art by the disclosure is such as to be commensurate with the scope of protection sought by the claims.").

143 *See* The Incandescent Lamp Patent, 159 U.S. 465 (1895); Robert P. Merges & Richard R. Nelson, *On the Complex Economics of Patent Scope*, 90 COLUM. L. REV. 839, 849–50 (1990) (discussing the case and the appropriate scope of enablement).

144 The possibility of regulatory pre-approval for black-box medicine algorithms has the potential to change this dynamic significantly. If a particular algorithm must be approved by a regulator—most likely, FDA—before marketing and use, inventing around becomes a less attractive option because alternate algorithms must also undergo a presumably costly FDA approval process. This patent-strengthening dynamic is seen in drug patents, which are especially strong because FDA approves exactly the drug covered (presumably) in the relevant patent, and not similar drugs. *See* W. Nicholson Price II, *Regulating Secrecy*, 91 WASH. L. REV. (forthcoming 2016) (discussing this dynamic). FDA approval would similarly strengthen

2. Written Description

The specification of a patent "shall contain a written description of the invention."[145] The Federal Circuit has held that this requirement is separate from the enablement requirement.[146] It serves to protect against overbroad or premature patenting, and limits the scope of the invention.[147] The inventor must actually possess the claimed invention at the time of filing—whether through actual or constructive reduction to practice—and must demonstrate that possession by describing the invention fully.[148]

Meeting the written description requirement ranges from moderately difficult, for instances where an algorithm is merely opaque-through-complexity, to extremely challenging, when the process is formally opaque.[149] In the first case, an algorithm can be stated even if the explanation offers little understanding; patents need not describe why the invention works, just that it does and how someone can replicate it.[150] In the second case, actually describing a fully opaque algorithm in words may be impossible. However, inventors may be able to demonstrate possession of the invention in the same way as they can enable inventions that require otherwise unavailable starting materials—by depositing everything necessary to recreate the algorithm (or the final algorithm, to the extent that is actually different) in a publicly accessible repository.[151] Another option, which has been suggested in the context of difficult-to-characterize biologic drugs, is making greater use of product-by-process claims, under which the

narrow black-box algorithm patents. Regulation of black-box medicine and medical algorithms generally is beyond the scope of this piece, but is discussed further in Price, *Black-Box Medicine*, *supra* note 7, and W. Nicholson Price II, *Regulating Algorithmic and Black-Box Medicine* 11–45 (Sept. 8, 2015) (unpublished manuscript) (on file with author).

145 35 U.S.C. § 112(a) (2012).

146 The written description requirement is analytically and practically distinct from enablement, though the two often stand or fall together, especially outside the biopharmaceutical industry. *See* Ariad Pharm., Inc. v. Eli Lilly & Co., 598 F.3d 1336, 1352–53 (Fed. Cir. 2010) (noting the analytical distinction between the two requirements and giving examples of enablement without adequate written description in biological and chemical contexts).

147 *Id.* at 1351 ("[T]he test for sufficiency is whether the disclosure of the application relied upon reasonably conveys to those skilled in the art that the inventor had possession of the claimed subject matter as of the filing date."). *See generally* Holbrook, *supra* note 133.

148 *Id.*

149 Price, *Describing Black-Box Medicine*, *supra* note 131, at 353–56.

150 *See* Newman v. Quigg, 877 F.2d 1575, 1581 (Fed. Cir. 1989) ("[I]t is not a requirement of patentability that an inventor correctly set forth, or even know, how or why the invention works . . . ."). Notably, however, with a lack of understanding, it becomes more difficult for the patentee to claim broader subject matter protection under the enablement requirement, as described above. *See supra* notes 142–44 and accompanying text.

151 *See* Price, *Describing Black-Box Medicine*, *supra* note 131, at 353–56.

algorithm is claimed via a detailed description of how it was generated, even if the algorithm itself cannot be fully described.[152]

3. Definiteness

Finally, section 112 requires that an invention be claimed in definite terms.[153] This requirement is intended to make clear to others the meets and bounds of the claimed invention.[154] Claims using ambiguous language, such as "words of degree," may be indefinite unless the context of the invention, the knowledge of the PHOSITA, and the disclosure of the specification, taken together, inform the PHOSITA of what is claimed with reasonable certainty.[155] Although definiteness has long been a very low bar for a patent applicant to meet, the Supreme Court recently took steps to raise that bar in *Nautilus v. Biosig Instruments*.[156] The Federal Circuit's reaction has been muted at best, leaving this area of law contested and unsettled.[157]

The ability of opaque algorithms to meet the evolving definiteness standard is currently untested. Presumably, courts will recognize the inherent indefiniteness in the field, which can make otherwise indefinite language permissible. On the other hand, the need to make terms as definite as possible will likely restrict the scope of what can be validly claimed, leading to narrow patents.[158]

---

152 *See* Dmitry Karshtedt, *Limits on Hard-to-Reproduce Inventions: Process Elements and Biotechnology's Compliance with the Enablement Requirement*, 3 HASTINGS SCI. & TECH. L.J. 109 (2011) (describing product-by-process claims for biotechnology inventions). Note that such deposits may face other challenges for black-box medicine algorithms, since reproducing them will typically require access to the same data as the original algorithm. However, making that data publicly available raises incentive problems for the first inventor and possibly other regulatory complications such as compliance with health privacy laws. As I argue here, *infra* Section IV.A, and elsewhere, Price, *Black-Box Medicine*, *supra* note 7, at 450, moving to a public or public/private infrastructure model with greater data disclosure would help resolve these problems.

153 35 U.S.C. § 112 (2012).

154 *See* Nautilus, Inc. v. Biosig Instruments, Inc., 134 S. Ct. 2120, 2124 (2014) ("[A] patent is invalid for indefiniteness if its claims, read in light of the specification delineating the patent, and the prosecution history, fail to inform, with reasonable certainty, those skilled in the art about the scope of the invention.").

155 Biosig Instruments v. Nautilus, Inc., 783 F.3d 1374, 1379–81 (Fed. Cir. 2015).

156 *See Nautilus*, 134 S. Ct. at 2120.

157 *See* Jason Rantanen, Teva*,* Nautilus*, and Change Without Change*, 18 STAN. TECH. L. REV. 430, 439–443 (2015) (describing the history of the definiteness requirement and the Supreme Court's holding in *Nautilus*, and noting that "[t]he Federal Circuit's post-*Nautilus* decisions do not even hint at a raised standard, either formally or in application").

158 *See supra* notes 142–44 and accompanying text. In section 112(f) means-plus-function claims, the Federal Circuit has held that algorithms (though not necessarily computer code) must typically be disclosed in the specifications to avoid patent invalidity for indefiniteness. *See, e.g.*, Blackboard, Inc. v. Desire2Learn Inc., 574 F.3d 1371, 1383–84 (Fed. Cir. 2009) (noting that

Overall, patent incentives are very low for black-box medicine. Algorithms are frequently unpatentable under section 101 doctrine, and in those instances where algorithms can constitute patentable subject matter—for instance, when they are linked with something substantially more than just the algorithm or law of nature itself—the requirements of section 112 will strongly limit the scope and availability of the resulting claims. Even when patents are available, they may be particularly difficult to enforce, as algorithm use may not be a visible part of medical care in many instances.[159] Patents provide little incentive for black-box algorithms, and none at all for investment in the hurdles of datasets or validation. Patents are not the only form of external incentive for investment in innovative information goods, however, and the next Part turns to non-patent incentives, considering both secrecy and other direct investment incentives.

III. NON-PATENT INCENTIVES

Other policy incentives can complement patent law. This Part describes the current incentive landscape for two types of innovation incentive: trade secrecy, which provides a limited form of government-protected appropriability; and direct government incentives, including grants, prizes, and tax incentives.

A. *Secrecy*

Trade secrecy is the principal private law alternative to the patent system for protecting technological innovation.[160] Actual secrecy—that is, effectively keeping information from other parties, including competitors—allows appropriation of innovative information and thus enables the innovator to charge supracompetitive prices. Actual secrecy is enhanced by the doctrine of trade secrecy.[161] Knowledge which is

---

an abstract "black box that performs a recited function" mentioned in the specification failed to disclose a structure sufficient for section 112(f), although an ordinarily skilled artisan might nonetheless have been enabled to practice the invention). Such disclosure runs into the linked problems mentioned above. *See supra* Sections II.B.1–2.

159 *See* Kapczynski & Syed, *supra* note 102 (noting that medical treatment knowledge, especially knowledge about what treatments don't work, is hard to exclude and that related patents are hard to enforce).

160 *See* Bone, *supra* note 83, at 243

161 *Cf.* Mark A. Lemley, *The Surprising Virtues of Treating Trade Secrets as IP Rights*, 61 STAN. L. REV. 311 (2008) [hereinafter Lemley, *The Surprising Virtues of Treating Trade Secrets as IP Rights*] (arguing that trade secrecy doctrine facilitates disclosure by reducing the need to invest in means to ensure actual secrecy).

reasonably kept secret and which derives independent economic value from its secrecy is protected from misappropriation by state and federal trade secret law.[162] However, secret information can legally be reverse-engineered or independently created.[163] Trade secrecy can protect information that is unpatentable, and lasts as long as the information is secret.[164]

Secrecy creates a mixed set of incentives and effects for black-box medicine. For two types of innovative information, datasets and algorithms, secrecy can protect effectively against competition, at least in some circumstances. On the other hand, secrecy also creates substantial problems for those two types of innovation by restricting both cumulative innovation and the benefits gained from others' access to aggregated information and algorithms. The third form of necessary innovation investment—validation—gains no incentives from secrecy, as validation must be shared to be valuable, but is hampered by secrecy applied to datasets or algorithms.

1. Data

Sets of data are not generally protectable with patents or copyrights in the United States, but can be kept secret.[165] The clearest example is Myriad Genetics itself.[166] Both before and after its loss in the Supreme Court, the company has kept much of its information about genetic variation secret.[167] Myriad's gene testing process reveals combinations of alleles present in patients; the company then offers free testing to family members, and analyzes family variation to determine significantly

---

162 Forty-seven states have enacted some form of the Uniform Trade Secrets Act, with the exception of North Carolina, New York, and Massachusetts, the latter two of which have planned 2016 introductions. *See Legislative Fact Sheet—Trade Secrets Act*, UNIFORM L. COMMISSION, http://www.uniformlaws.org/LegislativeFactSheet.aspx?title=Trade%20Secrets%20Act (last visited Jan. 14, 2016). Under federal law, the Economic Espionage Act of 1996 makes the theft or misappropriation of a trade secret a federal crime. *See* 18 U.S.C. § 1832 (2012).

163 *See* RESTATEMENT (THIRD) OF UNFAIR COMPETITION § 43 (AM. LAW INST. 1995) (listing reverse engineering as a proper means of acquiring a trade secret).

164 *See* Bone, *supra* note 83, at 248.

165 *See, e.g.*, Feist Publ'ns, Inc. v. Rural Tel. Serv. Co., 499 U.S. 340, 354 (1991) ("That there can be no valid copyright in facts is universally understood.") In Europe, a sui generis system of database protection has existed since 1996. Council Directive 96/9/EC, art. 3, 7, 1996 O.J. (L 077) 20 (EU), http://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:31996L0009&from=EN.

166 *See* Evans, *Economic Regulation of Next-Generation Sequencing*, *supra* note 29; *see also* Burk, *supra* note 61 (discussing the role of patents in allowing Myriad Genetics to assemble a large and useful dataset of BRCA1 and BRCA2 (breast and ovarian predisposition genes)).

167 Robert Cook-Deegan et al., *The Next Controversy in Genetic Testing: Clinical Data as Trade Secrets?*, 21 EUROPEAN J. HUM. GENETICS 585, 585–86 (2013).

linked genetic patterns.[168] Since Myriad has a substantially greater set of data on BRCA1/2 variants, only three percent of its samples have variants of unknown significance;[169] for competitors, the rate is roughly twenty percent to thirty percent.[170] Test samples sent to Myriad are, therefore, much less likely to be returned to the physician as "uninterpretable" than samples sent to its competitors,[171] providing a robust competitive advantage. While Myriad's data advantage could be overcome as other firms slowly assemble their own databases, the fact that Myriad currently possesses a much larger database—amassed from its period of patent protection—is self-reinforcing.[172] Myriad can provide more results, and is therefore likely to continue receiving more test samples, while the resulting larger database would still be kept as a trade secret.[173] Myriad's business plan includes retaining and expanding this secrecy-based advantage of mutation data and its relatively simple algorithms.[174]

Although the proprietary nature of datasets helps firms protect and recoup their investments, it also hampers future innovation.[175] This effect is especially important for black-box medicine because it relies on

---

168 *Id.* at 585.

169 *Id.* In a genetic test like Myriad's, the physical process first determines which alleles of a gene the patient has. That identification must then be interpreted to convey useful medical information: Are the alleles associated with a higher or a lower risk of cancer, or with no change? Douglas F. Easton et al., *A Systematic Genetic Assessment of 1,433 Sequence Variants of Unknown Clinical Significance in the BRCA1 and BRCA2 Breast Cancer-Predisposition Genes*, 81 AM. J. HUM. GENETICS 873, 873 (2007). When the interpreting entity lacks sufficient information about a particular allele to provide a useful interpretation, it is termed a "variant of unknown significance," and that part of the test is inconclusive. Cook-Deegan et al., *supra* note 167, at 585.

170 Cook-Deegan et al., *supra* note 167, at 585.

171 *Id.* at 585–86.

172 *See* Burk, *supra* note 61, at 239–40.

173 *Id.*

174 *Id.* at 240–41.

175 Keeping data proprietary raises several other potential concerns. On the ethical side, the Chairwoman of the European Society of Human Genetics' Professional and Public Policy committee described the committee as "very concerned that such important data is being withheld from those who most need it." Press Release, Eur. Soc'y of Human Genetics, Privately Owned Genetic Databases May Hinder Diagnosis and Bar the Way to the Arrival of Personalised Medicine: ESHG Reacts to Today's Report in the European Journal of Human Genetics (Oct. 31, 2012), https://www.eshg.org/13.0.html. She suggested that "[p]olicymakers should take an urgent look at the regulatory and reimbursement issues involved in genomic testing in order for all the data that is essential to understanding the clinical significance of [mutations] to be made public, to the benefit of patients and healthcare providers alike." *Id.* Others have noted that keeping data proprietary removes them from the potential of peer review and makes us less certain of their accuracy. *See* Misha Angrist & Robert Cook-Deegan, *Distributing the Future: The Weak Justifications for Keeping Human Genomic Databases Secret and the Challenges and Opportunities in Reverse Engineering Them*, 3 APPLIED & TRANSLATIONAL GENOMICS 124, 125 (2014). Other concerns arise with respect to transparency, oversight, and the blocking of future research directions.

very large datasets to identify and use otherwise inaccessible complex patterns. When data are kept in company-specific silos that are focused on a particular set of identifiable characteristics or relationships, those data are typically unavailable for use in finding relationships that we have not yet—or cannot yet—identify. Thus, even if proprietary datasets can be justified, arguendo, to facilitate specific, explicit forms of innovative medicine,[176] that justification may fall in the context of developing next-generation algorithms such as those in black-box medicine. Trade secrecy slows cumulative innovation and promotes duplicative investment,[177] and these dynamics are especially strong in this context.

2. Algorithms

Secrecy for algorithms also provides incentives by preventing appropriation, but creates a different set of problems. A black-box algorithm itself is unavoidably secret, by definition, but the surrounding information—principally, how the algorithm is developed and trained—can be disclosed or kept secret. If the bulk of information about an algorithm is kept secret, only the inventor can use it, and can accordingly charge monopoly rents (up to the value of the product, of course). Secrecy is accordingly used frequently to protect the commercial value of algorithms in other fields, and could be similarly used for black-box medicine.[178]

The challenge is that secret algorithms are hard to oversee, hard to trust, hard to validate, and cannot form the foundation of later cumulative innovation. Secret algorithms would be more difficult for regulators to oversee, although mechanisms do exist for regulators like FDA to receive and evaluate secret information while maintaining its confidentiality.[179] These mechanisms are harder to stretch to private third parties who would need information about how an algorithm was developed in order to verify its quality. From a policy perspective,

---

176 *See* Burk, *supra* note 61; Evans, *Economic Regulation of Next-Generation Sequencing*, *supra* note 29, at 56.

177 Bone, *supra* note 83, at 266, 269; Robert G. Bone, *The (Still) Shaky Foundations of Trade Secret Law*, 92 TEX. L. REV. 1803 (2014). For a defense of treating trade secrecy as intellectual property, see Lemley, *The Surprising Virtues of Treating Trade Secrets as IP Rights*, *supra* note 161, at 329–38.

178 In *The Black-Box Society*, Frank Pasquale extensively describes the role of deliberately opaque algorithms, which he dubs "black-box" algorithms based on that deliberate opacity, and cites the extensive commercial use of secret algorithms in, inter alia, the financial industry. *See* PASQUALE, *supra* note 11.

179 *See, e.g.*, 21 U.S.C. § 331(j) (2012) (prohibiting the use of information concerning "any method or process which as a trade secret is entitled to protection" that is submitted to FDA).

secrecy about algorithm development makes cumulative innovation harder, as others cannot learn what works and what doesn't. Finally, deliberately secret algorithms amplify the challenge of getting doctors and patients to adopt medical technology that is not well-understood and that is, in fact, not fully understandable.[180]

3. Validation

Finally, as described above, third-party validation of black-box algorithms is key for demonstrating their accuracy and utility. Therefore, secrecy is unavailable to create incentives for third parties to validate the strength and accuracy of algorithms. Secrecy of algorithms themselves also makes validation harder by limiting third-party access to the details needed to validate those algorithms.

B. *Non-Exclusivity Incentives*

In addition to the exclusivity incentives described above—patents and trade secrecy—other policy incentives are available to spur innovation. A significant and expanding scholarly literature addresses the use of grants and prizes as innovation incentives,[181] and tax incentives for research have recently been recognized as an important part of the mix.[182] These mechanisms differ from exclusivity incentives in that rather than relying on appropriation or exclusivity to allow supracompetitive pricing, which recompenses innovation ex post through higher prices on the users of an innovation, they typically provide innovators with funds raised from a broader taxpayer base (whether ex ante or ex post).[183]

Because these funding mechanisms do not rely on exclusivity, they do not themselves impose limits on others' use of the innovation, which can potentially increase distribution of the innovation. Non-exclusivity mechanisms can, however, be combined with exclusivity mechanisms, and frequently are; an invention may be funded by a federal grant in a research university, patented by the university, and then licensed to a

---

180 *See* Price, *Black-Box Medicine*, *supra* note 7, at 465.

181 *See, e.g.*, Michael Abramowicz, *Perfecting Patent Prizes*, 56 VAND. L. REV. 115 (2003); Daniel J. Hemel & Lisa Larrimore Ouellette, *Beyond the Patents-Prizes Debate*, 92 TEX. L. REV. 303 (2013); Benjamin N. Roin, *Intellectual Property Versus Prizes: Reframing the Debate*, 81 U. CHI. L. REV. 999 (2014) [hereinafter Roin, *Intellectual Property Versus Prizes*]; Joseph Stiglitz, *Give Prizes Not Patents*, NEW SCIENTIST, Sept. 16, 2006, at 21.

182 Hemel & Ouellette, *supra* note 181.

183 *Id.*

firm that receives tax incentives for further research before commercial production.[184] It is thus a mistake to think that incentive mechanisms that do not require exclusivity will result in wide access to the innovation. If it is desirable that exclusivity be avoided, non-exclusivity funding mechanisms need to contain conditions to that effect, such as a prize requiring that competing firms commit not to patent the invention.

This Section does not attempt to catalog the prizes, grants, and tax incentives that may be generally available for black-box medicine. Since the field is burgeoning, specifically focused incentives are substantially rarer than general-purpose incentives. There is, however, one notable exception: President Obama's Precision Medicine Initiative.[185]

The Precision Medicine Initiative, announced in the State of the Union Address in 2015, proposes to direct $215 million to personalized medicine research over the next ten years.[186] In the short term, the initiative focuses on more explicit personalized medicine goals, particularly identifying differentiable cancer treatments in partnership with pharmaceutical companies and through typical clinical trial methods.[187] However, the longer-term goals of the initiative are (1) to support a national scientific network and (2) to develop a "national cohort study" of at least one million participants who will share genomic information and biological samples as well as clinical health data.[188] The former goal presumably includes additional grant funding targeted at personalized medicine, including black-box methods; the latter goal is an important step in addressing the data hurdle,[189] and recognizes the

184 For instance, researchers at Columbia University developed cotransformation, a powerful and basic biotechnology process for introducing foreign DNA into eukaryotic cells. *See* Alessandra Colaianni & Robert Cook-Deegan, *Columbia University's Axel Patents: Technology Transfer and Implications for the Bayh-Dole Act*, 87 MILBANK Q. 683 (2009) (describing the development of the technology). The initial work was funded by grants from the National Institutes of Health and the National Science Foundation. Michael Wigler et al., *Transformation of Mammalian Cells with Genes from Procaryotes and Eucaryotes*, 16 CELL 777, 785 (1979). Patents on the resulting technology were licensed to thirty-four firms and eventual licensing revenues reached $790 million. Colaianni & Cook-Deegan, *supra*, at 700. Biogen licensed the patented technology in its Multiple Sclerosis drug Avonex. Ted Agres, *Columbia Patents Under Attack*, SCIENTIST (July 25, 2003), http://www.the-scientist.com/?articles.view/articleNo/22353. The company benefits from substantial tax credits for its research and development spending. *See Taking Credit*, WALL ST. J. (June 14, 2013), http://www.wsj.com/articles/SB10001424127887324688404578543791391933684 (listing corporate R&D tax benefits).

185 *See* Fact Sheet: President Obama's Precision Medicine Initiative, *supra* note 34.

186 *Id.*

187 *See* Collins & Varmus, *supra* note 4, at 793–94; *see also Precision Medicine Initiative: Near-Term Goals*, NAT'L INST. HEALTH, http://nihprod.cit.nih.gov/precisionmedicine/goals.htm (last updated Oct. 13, 2015).

188 *See* Collins & Varmus, *supra* note 4, at 794–95.

189 *See* discussion *supra* Section I.B.1.

significant upside of centralized data collection for distributed research.[190]

The Precision Medicine Initiative is a significant step forward in non-exclusivity incentives and aids to innovation. However, the innovation incentive story remains problematic at an overall level, with relatively few incentives available for the development of black-box medicine.

IV. IMPROVING INCENTIVES

The absence of appropriate incentives significantly impedes the development of black-box medicine; more directly, some available incentives, like the particular contours of patentable subject matter, actively drive the development of personalized medicine in unhelpful directions.[191] Accordingly, the final Part of this Article briefly proposes potential improvements to the existing incentive structure.

Because the development process involves distinct forms of innovation, incentives are most usefully considered separately for each form. Generating large and well-curated datasets likely requires the greatest investment. Patents are unavailable for datasets, and trade secrecy is relatively ill-suited to consolidation and cumulative innovation. Instead, the amassing of high-quality datasets might better be conceived as an infrastructure for further innovation, which suggests a role for more direct government involvement. An additional possibility is the implementation of a tailored sui generis dataset protection regime, such as exists under European Union law.[192]

Second, black-box medicine development requires incentives for the generation of algorithms. As described above, algorithms were previously patent-eligible,[193] so one potential incentive for developing algorithms would come from reinstating patent protection for them. This solution, however, comes with its own set of complications. Regulatory exclusivity might be preferable, although that would require a regulatory preapproval regime that currently does not exist. Prizes are another potential solution[194]; although they are subject to many of the same general innovation considerations as patents, they are typically more flexible to implement.

Third, incentives are needed for validation. An ever-present concern in complex implicit models—especially when developed via

190 *See infra* Section IV.A.

191 *See* discussion *supra* Section II.A.3.

192 *See* Council Directive 96/9/EC, art. 3, 7, 1996 O.J. (L 077) 20 (EU).

193 *See* discussion *supra* Section II.A.1.

194 *See* discussion *supra* Section III.B.

black-box methods—is ensuring that they are valid and generally applicable, rather than just statistical artifacts arising from over-specification in large datasets. This Article will thus present potential structures for regulatory "bounties": rewards provided to competitor firms for either validating or substantially falsifying the black-box medicine algorithms of the innovator firms.

A. *Incentives for Datasets: The Infrastructure Model*

As described above, significant hurdles exist in the collection of large, high quality datasets available for the development of black-box medicine algorithms.[195] Patents are unavailable and trade secrecy presents problems discussed above: it lends itself to fragmenting rather than consolidating information, restricts cumulative innovation, and creates advantages for incumbents—like Myriad Genetics—which may continue indefinite specific monopolies.[196] To increase incentives, therefore, policymakers could turn to direct government intervention or a public-private partnership focused on data as infrastructure.[197]

In the context of genetic testing, the secrecy that protects the databases of incumbent firms has been analogized to an infrastructure problem, wherein specific sets of correlations—namely, the significance of individual genetic variations—have several features of essential facilities.[198] Datasets for black-box medicine development may similarly take the role of common infrastructure for further innovation. Rather than conceiving of datasets as innovation themselves, they could be viewed as shared resources that enable firms to develop innovative

---

195 *See infra* Section II.A.

196 For a detailed description of this problem, see Evans, *Economic Regulation of Next-Generation Sequencing*, *supra* note 29.

197 *See* Brett M. Frischmann, *An Economic Theory of Infrastructure and Commons Management*, 89 MINN. L. REV. 917 (2005) (discussing infrastructure and the benefits of managing infrastructure as a commons when social and public outputs are facilitated). Alternately, policymakers could attempt to develop a sui generis system of database protection modeled on the European system, though that would fail to solve the problems of data fragmentation. Under Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the Legal Protection of Databases, databases, defined as "a collection of independent works, data or other materials arranged in a systematic or methodical way and individually accessible by electronic or other means," Council Directive 96/9/EC, art. 1, 1996 O.J. (L 077) 20 (EU), are covered by copyright if they involve creative choices, but are otherwise covered by a sui generis intellectual property right, so long as their creation involved "substantial investment." *Id.* art. 3, 7. Database owners may prevent others from extracting or re-utilizing the whole or a substantial part of the database for approximately fifteen years from publication or, if the database is kept private, completion, and may be extended by additional substantial investment. *Id.* art. 10.

198 Evans, *Sustainable Access to Data*, *supra* note 54.

algorithms.[199] As Brett Frischmann has noted, accessible infrastructure resources are particularly valuable where social and public gains may be difficult to value commercially; broad improvements in medical care fit this model well.[200]

Under this view, direct or indirect government intervention could usefully aid the generation of datasets. On a direct level, collecting data shows a prima facie advantage for government. In the United States, many millions of patients participate in Medicare and Medicaid, where the government provision of insurance allows access to patients' medical records.[201] The Department of Defense and the Department of Veterans Affairs provide direct health care to and consequently collect data for over eleven million military personnel, veterans, and their families.[202] In other nations, the concentration of data with the government is even stronger. For example, the U.K. National Health Service (NHS) provides free health care to over sixty-three million U.K. residents,[203] and consequently accumulates tremendous amounts of data.

Government possession of data brings its own challenges. For instance, it is emphatically *not* the case that the data gathered by the Department of Defense and Department of Veterans Affairs are neatly available in high-quality interoperable formats. In fact, the two agencies have spent billions trying and failing to upgrade their electronic records systems, which remain incompatible.[204] And the U.K.'s NHS, while it has a great deal of data, is prevented by strict privacy rules from using

---

199 Others have called for collation and availability of health data, notably Marc A. Rodwin, *The Case for Public Ownership of Patient Data*, 302 J. AM. MED. ASS'N 86 (2009); *cf.* Evans, *Sustainable Access to Data*, *supra* note 54 (discussing data access to FDA's Sentinel data); Evans, *Economic Regulation of Next-Generation Sequencing*, *supra* note 29 (describing concerns about data sharing of genomic information).

200 Frischmann, *supra* note 197, at 996 (using malaria research as an example of an endeavor with high social value, but with low commercial value relative to the research costs).

201 *See Total Monthly Medicaid and CHIP Enrollment*, KAISER FAMILY FOUND., http://kff.org/health-reform/state-indicator/total-monthly-medicaid-and-chip-enrollment (last visited Jan. 14, 2016) (noting 71.8 million Medicaid/CHIP enrollees in October 2015); *Total Number of Medicare Beneficiaries*, KAISER FAMILY FOUND., http://kff.org/medicare/state-indicator/total-medicare-beneficiaries (last visited Jan. 14, 2016) (noting 49.4 million Medicare beneficiaries in 2012).

202 Tricare, which offers health for military personnel, had 9.2 million eligible beneficiaries in 2008. *See* TRICARE, U.S. DEP'T OF DEFENSE, BASIC FACTS OF THE MILITARY HEALTH SYSTEM (2008), http://www.tricare.mil/stakeholders/statistics.cfm. The Veterans Health Administration had 9.1 million enrollees in 2014. *See* VETERANS HEALTH ADMIN., U.S. DEP'T OF VETERANS AFFAIRS, SELECTED VETERANS HEALTH ADMINISTRATION CHARACTERISTICS: FY2002 TO FY2014 (2015), http://www.va.gov/vetdata/docs/Utilization/VHAStats_2014.xlsx.

203 *About the National Health Service (NHS)*, NAT'L HEALTH SERV., http://www.nhs.uk/NHSEngland/thenhs/about/Pages/overview.aspx (last updated July 1, 2015).

204 Hannah Winston, *Billions Wasted on Fruitless Bid to Create Paperless Vet Health Records*, NBC NEWS (Aug. 26, 2013, 3:32 AM), http://www.nbcnews.com/news/other/billions-wasted-fruitless-bid-create-paperless-vet-health-records-f8C11001233.

much of it, and certainly from sharing information with third-party private algorithm developers.[205] However, government entities are taking steps in the direction of data collection, even if not in sharing: the Veterans Health Administration is well into its effort to collect genetic and phenotypic information on a million veterans for research purposes.[206] But there are currently no indications that this information—or other information like it—will be made available for further innovation by private entities.

The other existing large government initiative in this area is FDA's Sentinel initiative. Sentinel is designed to collect and analyze drug safety data on very large populations—with a goal of acquiring health data on over 100 million Americans—to provide further information on post-approval drug safety.[207] In the Sentinel project, FDA has taken on the role of an infrastructure regulator in facilitating the development of a longitudinal health records database of insurance claims information, medical records, prescription drug records, and information from military care and Medicare.[208] FDA's involvement, along with the authorizing statute, enables FDA to cut through some of the legal and practical hurdles to assembling datasets; the drug safety analysis is conceived as part of the agency's public health authority, which allows it to fit into an exception to normal limitations on transferring and using identifiable data.[209] A pilot project, Mini-Sentinel, has already been developed in association with private institutional partners, and is actively engaged in drug safety research.[210] FDA's Sentinel-related safety-centered statutory authority could potentially extend to cover substantial swathes of research on differential impacts of drugs, since efficacy and safety are intertwined in evaluating drugs.[211] Furthermore,

---

205 Others have noted the richness of the NHS's data and the challenge of its privacy rules. *See* Wayne Parslow, *How Big Data Could Be Used to Predict a Patient's Future*, GUARDIAN (Jan. 17, 2014, 3:30 AM), http://www.theguardian.com/healthcare-network/2014/jan/17/big-data-nhs-predict-illness ("Although currently shielded by privacy rules, the personal data that can risk score every NHS patient already exists. And it is already far more centralised and normalised than in countries such as the US, giving the UK the opportunity to become the world leader.").

206 *See Million Veteran Program (MVP)*, U.S. DEP'T OF VETERANS AFF., http://www.research.va.gov/MVP (last updated Dec. 15, 2015).

207 *See* 21 U.S.C. § 355(k)(3)(B)(ii)(II) (2012) (setting a goal of 100 million patients for postmarket risk identification and safety analysis by July 1, 2012).

208 *Id.* § 355(k)(3)(C)(i)(III)(aa)–(cc); *see also* Barbara J. Evans, *Congress' New Infrastructural Model of Medical Privacy*, 84 NOTRE DAME L. REV. 585, 588 (2009) [hereinafter Evans, *Congress' New Infrastructural Model of Medical Privacy*].

209 Evans, *Congress' New Infrastructural Model of Medical Privacy*, *supra* note 208, at 598.

210 Susan Forrow et al., *The Organizational Structure and Governing Principles of the Food and Drug Administration's Mini-Sentinel Pilot Program*, 21 PHARMACOEPIDEMIOLOGY & DRUG SAFETY (SUPPLEMENT S1) 12 (2012).

211 Evans, *Congress' New Infrastructural Model of Medical Privacy*, *supra* note 208, at 601. On the other hand, broad research questions that stray beyond the actual public health use

FDA could centralize the data it collects and make the data available to other researchers to pursue these research avenues.[212] However, the Sentinel project is not taking this broad and centralized infrastructure path. Instead, so far Mini-Sentinel has been focused on traditional views of drug safety and adverse events.[213] More importantly, rather than creating a centralized dataset, FDA has pursued a distributed model in which it develops research queries and then transmits those queries to its institutional partners (hospitals, health systems, and the like) that actually hold the data, before those partners transmit de-identified answers back.[214] Thus, although Sentinel has the potential and promise to be the type of centralized research dataset that could provide wide-ranging infrastructure for the development of black-box medicine, it seems that FDA is taking the project in a different direction. Nevertheless, the structure of Sentinel, and the authority that is vested in FDA by its statute, indicates at least the potential for broader government intervention into enabling datasets for future research.

The federal government is also pursuing a different, de novo path to creating a new dataset as part of the Precision Medicine Initiative, which seeks to support personalized medicine, in part by creating a national cohort study containing the genomic data, clinical data, and biological samples of at least one million subjects;[215] this would be much smaller than the Sentinel dataset, but still quite substantial. Although the exact contours of the eventual cohort are unclear, it is likely to involve existing patient networks.[216] The Initiative also clearly contemplates federal funding to lay the groundwork for this dataset—the infrastructure for the data infrastructure, so to speak—including a projected ten million dollars for FDA "to acquire additional expertise and advance the development of high quality, curated databases to support the regulatory structure needed to advance innovation in precision medicine and protect public health," and five million dollars for the Office of the National Coordinator for Health Information

function of the Sentinel system might run afoul of HIPAA regulations that treat research differently. *Id.* at 615–16. Further public perception difficulties may arise if FDA broadened Sentinel's mission, since there is a distinct tradeoff between public health benefits and privacy; as research broadens, privacy protections diminish when more data is spread among more research projects and, presumably, more researchers. *Id.* at 605.

212 *Id.* at 601–03.

213 *See Background*, MINI-SENTINEL, http://mini-sentinel.org/about_us/default.aspx (last updated Sept. 18, 2014) ("Mini-Sentinel monitors the safety of FDA-regulated medical products through assessment of routinely collected electronic healthcare data in response to FDA concerns.").

214 Forrow et al., *supra* note 210, at 14–15.

215 Fact Sheet: President Obama's Precision Medicine Initiative, *supra* note 34.

216 *Id.* ("This ambitious project will leverage existing research and clinical networks and build on innovative research models that enable patients to be active participants and partners.").

Technology "to support the development of interoperability standards and requirements that address privacy and enable secure exchange of data across systems."[217] The data from this cohort study will be "broadly accessible to qualified researchers," though the definition of "qualified researchers" and any limitations to the acceptable research use of the data remain to be determined.[218]

The government could take this approach in a different direction to enable black-box medicine (and other personalized medicine) by simplifying the data collection step, generating the dataset-infrastructure, and then allowing private parties to compete in the analysis and validation steps. Data access could be leveraged in at least two ways. First, the data could be used exclusively for some time; firms could bid for access to segments of the data, coupled with a commitment to make any resulting algorithms publicly available after some period of exclusivity.[219] Second, the data could be made freely available, but with the caveat that firms using the data disclose their algorithms. This would enable a broader set of concurrent developments, while still allowing firms to capture benefits of their (reduced) innovation investments.[220]

Similar options exist to pursue database generation through public-private partnerships. deCODE Genomics famously exemplifies such a partnership. The Icelandic biopharmaceutical firm successfully lobbied the Icelandic Parliament to create a population-wide Health Sector Database including genomic, genealogical, and health information.[221] Court challenges shifted the database from mandatory to voluntary,[222] and the effort was highly controversial,[223] but over 160,000 individuals still volunteered,[224] and the company has published extensively on the

---

217 *Id.*

218 *See id.*

219 This approach has potential political economy problems related to its implementation; the specter of the government collecting health records and turning them over to private parties for their exclusive benefit would likely meet substantial resistance. This dynamic certainly exists in other frameworks—notably in the patenting of government-funded innovation under the Bayh-Dole Act, Pub. L. No. 96-517, sec. 6(a), §§ 200–211, 94 Stat. 3015, 3019—but may be even more politically sensitive in the context of health information.

220 This plan would also potentially avoid some of the political economy problems discussed *supra* note 219. However, it would exacerbate privacy and reidentification concerns. *See* Cohen et al., *supra* note 43.

221 Vilhjálmur Árnason, *Coding and Consent: Moral Challenges of the Database Project in Iceland*, 18 BIOETHICS 27 (2004).

222 Renate Gertz, *An Analysis of the Icelandic Supreme Court Judgement on the Health Sector Database Act*, 1 SCRIPT-ED 241 (2004).

223 *See, e.g.*, Árnason, *supra* note 221; Jeffrey R. Gulcher & Kari Stefánsson, *The Icelandic Healthcare Database and Informed Consent*, 342 NEW ENG. J. MED. 1827 (2000).

224 *Unrivaled Capabilities*, DECODE GENETICS, http://www.decode.com/research (last visited Feb. 1, 2016).

explicit genomic links it has found.[225] Similarly, the Human Genome Project provides another clear precedent. There, a collaborative effort between government and private researchers sequenced the human genome with the intention of providing it freely to future researchers and innovators as a common infrastructure resource.[226]

B. *Incentives for Algorithms*

The heart of black-box medicine is the development of biomedically useful algorithms by plumbing the masses of health data.[227] However, as described above, this process is neither easy nor inexpensive.[228] And current intellectual protection is both inadequate and skewed away from black-box medicine. Accordingly, better incentives are needed to drive algorithm development. Potential incentives could come in at least three forms: patent protection, regulatory exclusivity, or prizes.

1. Patents

Patents are an obvious source of incentives for algorithms, as they were generally available for algorithms until the Supreme Court's decision in *Prometheus*.[229] Congress could override that statutory interpretation decision by amending the statute to, for instance, state that novel diagnostic algorithms are patent-eligible subject matter. While this approach is initially attractive, challenges arise in both enactment and enforcement.

First, overruling *Prometheus* may face problems of overbreadth. In particular, black-box medicine is similar to mainstream computer software patents and algorithms, which are criticized by academics, frequently disliked by the software industry itself, and a target of reform efforts.[230] Broad-brush patent changes to revive algorithmic patents

---

225 *See, e.g.*, Thorlakur Jonsson et al., *A Mutation in APP Protects Against Alzheimer's Disease and Age-Related Cognitive Decline*, 488 NATURE 96 (2012); Unnur Styrkarsdottir et al., *Nonsense Mutation in the LGR4 Gene Is Associated with Several Human Diseases and Other Traits*, 497 NATURE 517 (2013).

226 Francis S. Collins et al., *The Human Genome Project: Lessons from Large-Scale Biology*, 300 SCI. 286 (2003).

227 *See* discussion *supra* Section I.A.

228 *See* discussion *supra* Section II.B.

229 *See* discussion *supra* Section II.A.1.

230 *See, e.g.*, Colleen V. Chien, *Reforming Software Patents*, 50 HOUS. L. REV. 325 (2012); Jay Dratler, Jr., *Does Lord Darcy Yet Live? The Case Against Software and Business-Method Patents*, 43 SANTA CLARA L. REV. 823 (2003); Pamela Samuelson, *Benson Revisited: The Case Against Patent Protection for Algorithms and Other Computer Program-Related Inventions*, 39 EMORY

may, therefore, face considerable resistance and may also have negative impacts on other industries.[231] Finally, of course, the Supreme Court may have correctly judged the innovation incentives regarding laws of nature and determined that patents on relationship-based algorithms may be harder to invent around and therefore may block overall innovation.[232]

Second, patents granted on black-box medicine algorithms face significant difficulties in enforcement. Knowing whether infringement is occurring and proving that it has occurred are both likely to be difficult, especially for more complex algorithms.[233] Thus, though restoring the patent system to its status before *Prometheus* has some initial appeal for driving the development of stable, well-defined algorithms, other possibilities may better align incentives to drive innovation forward.

2. Regulatory Exclusivity

Regulatory exclusivity could provide incentives better tailored to algorithms. Instead of relying on the patent system to provide the incentive of excludability, in regulatory exclusivity a regulator restricts competition by limiting the availability of pre-market approval to competitors.[234] Thus, regulatory exclusivity requires the existence of a

---

L.J. 1025 (1990); Robert E. Thomas, *Debugging Software Patents: Increasing Innovation and Reducing Uncertainty in the Judicial Reform of Software Patent Law*, 25 SANTA CLARA COMPUTER & HIGH TECH. L.J. 191 (2008). *But see* Martin Campbell-Kelly & Patrick Valduriez, *A Technical Critique of Fifty Software Patents*, 9 MARQ. INTELL. PROP. L. REV. 249 (2005) (finding the most frequently cited software patents worth protection).

231 Merely restoring patent law incentives for black-box medicine might also result in problems for black-box medicine itself. Although this Article begins with the position that *Prometheus* and *Myriad* problematically reduced the patent incentives available for black-box medicine, *see* discussion *supra* Section II.A.2, those initial incentives had flaws as well. In particular, since black-box medicine resembles software in many ways, we might expect to see some of the same problems that infect patents in the software industry, including frequent issuance of patents on obsolete technologies, a mismatch between patent claims and actually invented subject matter, broad and vague claim language, and substantial transaction costs. *See, e.g.*, Chien, *supra* note 230; Mark A. Lemley, Address, *Software Patents and the Return of Functional Claiming*, 2013 WIS. L. REV. 905 (2013); Thomas, *supra* note 230.

232 The scholarly debate on these issues is extensive and need not be recapped here. *See* Mark A. Lemley et al., *Life After* Bilski, 63 STAN. L. REV. 1315 (2011). For a discussion of the relevant knowledge/embodiment distinction in patent law, see Kevin Emerson Collins, *The Knowledge/Embodiment Dichotomy*, 47 U.C. DAVIS L. REV. 1279 (2014).

233 For an analogous situation, see, for example, W. Nicholson Price II, *Making Do in Making Drugs: Innovation Policy and Pharmaceutical Manufacturing*, 55 B.C. L. REV. 491, 526–27 (2014) [hereinafter Price, *Making Do in Making Drugs*] (describing the difficulties secrecy creates in enforcing manufacturing process patents).

234 *See, e.g.*, Eisenberg, *The Role of the FDA in Innovation Policy*, *supra* note 71; Yaniv Heled, *Regulatory Competitive Shelters*, 76 OHIO ST. L.J. 299 (2015). "Exclusivity," though widely used,

premarket approval regime. In multiple contexts where such preapproval requirements exist, some form of regulatory exclusivity is used as an innovation incentive;[235] in others, it has been proposed.[236] FDA administers the majority of extant applications of regulatory exclusivity, primarily around the marketing of small-molecule drugs and biologics.[237]

Assuming the existence of a preapproval regime, regulatory exclusivity could function the same way for black-box medicine predictive models as for other innovations within preapproval regimes. For example, in the context of drugs, FDA will not approve a generic drug within five years of the approval of a drug based on a new chemical entity.[238] For biologics, the period is twelve years.[239] Similarly, if FDA approval were required for black-box medicine models to be commercially marketed and used, FDA could withhold that approval from imitator products for a fixed period of time as a reward to the innovator company.

The main advantages for regulatory exclusivity are flexibility, ease of enforcement, and strong disclosure. Regulatory exclusivity is more flexible than the patent system. It is administered by an expert agency with experience in the specific technology and—ideally—an innovation mandate as well as a regulatory health and safety mandate.[240] Even without substantial changes to the statutory contours of exclusivity, the agency can apply it flexibly. Statutory changes are also easier because regulatory exclusivity is not bound by the same treaty requirements as

---

is somewhat inaccurate because competitors are not excluded from the market; instead, a faster, cheaper path to market—whether biosimilar approval or an Abbreviated New Drug Application—is foreclosed for some period, effectively sharply increasing the costs of approval and practically limiting market entry. *Id.* at 318–19. Heled proposes the more general term "regulatory competitive shelter" to describe this phenomenon, *id.*, but I will continue to use "regulatory exclusivity," largely for simplicity's sake. To the extent that regulatory competitive shelters could take on more shades than pure exclusivity—for instance, purely higher costs to market entry, in the nature of regulator enforced mandatory license fees—the broader phenomenon might more appropriately describe alternate solutions.

235 *See* Heled, *supra* note 234 (describing regulatory exclusivity regimes for drug, biologic, and pesticide development).

236 *See* Price, *Making Do in Making Drugs*, *supra* note 233 (proposing regulatory exclusivity for drug manufacturing innovations to promote such innovation).

237 Heled, *supra* note 234 (listing fourteen such regimes, of which thirteen are administered by FDA and one by the Environmental Protection Agency).

238 21 U.S.C. § 355(j)(5)(F)(ii) (2012) (granting five years of market exclusivity for new chemical entities).

239 *See* 42 U.S.C. § 262(k)(7) (2012) (granting four years of market exclusivity and an additional eight years of data exclusivity to biologics).

240 Eisenberg, *The Role of the FDA in Innovation Policy*, *supra* note 71.

patent law; it can be flexible across products and across industrial sectors in a way that patent law cannot.[241]

The second advantage is that regulatory exclusivity is substantially easier to enforce than patents, with consequently more uniform enforcement. The default of a market preapproval regime is the inability to enter the market; thus, if regulatory exclusivity exists for a particular well-defined product, competitors can be prevented from entering that market simply by denying approval for the competitors' products for the appropriate period of time.[242] This contrasts with the difficulty and expense of enforcing patents.[243]

The third and final advantage to regulatory exclusivity comes only if exclusivity is coupled with a disclosure requirement. In the context of drug development, regulatory exclusivity demands the production of knowledge (that a drug is safe and effective, as measured by clinical trials), and requires at least some disclosure of that knowledge.[244] Although clinical trial data are not fully disclosed now,[245] the basic results of trials—that a particular drug is safe and effective for a particular indication—become public and can eventually be relied upon

---

241 The treaty on Trade Related Aspects of Intellectual Property Rights (TRIPS), to which the United States is a party, requires that patent systems be relatively uniform across different countries. *See* Agreement on Trade-Related Aspects of Intellectual Property Rights, Apr. 15, 1994, Marrakesh Agreement Establishing the World Trade Organization, Annex 1C, 108 Stat. 4809, 1869 U.N.T.S. 299. For the purposes of this Article, the most important requirement of TRIPS is that patent terms cannot be technology-specific. Eisenberg, *The Role of the FDA in Innovation Policy*, *supra* note 71, at 365 (suggesting that regulatory exclusivity may help tailor innovation policy without violating TRIPS). Note, however, that countries may choose to disallow the patentability of medical techniques, as Europe has largely done, and as the United States has effectively done. *See* EPC art. 52(2), *supra* note 130; *see also* 35 U.S.C. § 287(c) (2012). Black-box medicine and algorithmic medicine in general could potentially be excluded from patentability, but probably not given an intermediate or differently structured set of incentives from other technological areas.

242 On the other hand, the definition of a "product" might be particularly flexible in the context of black-box medicine; fluid boundaries would raise some of the same enforcement challenges that exist in patent law.

243 *See* AM. INTELLECTUAL PROP. LAW ASS'N, REPORT OF THE ECONOMIC SURVEY: 2011 (noting that, for patent infringement claims under $1 million, median legal costs are $650,000; for claims from $1 million to $25 million, costs are $2.5 million; for claims over $25 million, median costs are $5 million).

244 *See* Eisenberg, *The Role of the FDA in Innovation Policy*, *supra* note 71, at 366–72.

245 A rapidly growing movement focuses on the disclosure of clinical trial data. *See* Kamran Abbasi, *Compulsory Registration of Clinical Trials*, 329 BMJ 637 (2004); Richard Lehman & Elizabeth Loder, *Missing Clinical Trial Data*, 344 BMJ d8158 (2012); Michelle M. Mello et al., *Preparing for Responsible Sharing of Clinical Trial Data*, 369 NEW ENG. J. MED. 1651 (2013). However, these disclosures are not tied to FDA approval. Arguments have long been made that information submitted for regulatory approval should be disclosed, *see* Thomas O. McGarity & Sidney A. Shapiro, *The Trade Secret Status of Health and Safety Testing Information: Reforming Agency Disclosure Policies*, 93 HARV. L. REV. 837 (1980), but those arguments have not succeeded.

by generic companies to secure approval.[246] In general, since the regulator who approves the product is the same entity that enforces regulatory exclusivity, innovators have an incentive to be forthcoming and candid in their disclosures, rather than facing the incentive to obscure useful technical information in patents to minimize disclosure to competitors.[247]

The principal challenge with implementing regulatory exclusivity is that it relies on a market-spanning regulatory preapproval regime, which does not currently exist for data-driven diagnostic tests.[248] A full analysis of FDA's diagnostic test regime and what is most appropriate for black-box medicine must await future work. In brief, however, while FDA does currently regulate some diagnostic tests, many exist outside its current scope, and there is certainly not a comprehensive regime in place.[249] Were such a regime implemented, regulatory exclusivity would be an attractive possibility. Nevertheless, the possibility of regulatory exclusivity as an innovation incentive is probably not sufficient justification to impose a premarket approval regime if one would otherwise not be warranted.

Other problems with regulatory exclusivity are inherent in the name and the concept: it, like the patent system, focuses on exclusivity. To the extent that black-box medicine models rely on underlying natural laws, excluding others from using those laws presents the same preemption problems that the Supreme Court named as problematic for innovation in *Prometheus*.[250] Additionally—and problematically—applying regulatory exclusivity relies on defining the contours of a specific model. When models are multifaceted, complex, and implicit, defining the contours of a model and knowing whether another model overlaps with those contours may be an insurmountable hurdle.[251]

---

246 21 U.S.C. § 355(b)(2) (2012) (allowing generic drug applicants to rely on the finding of safety and efficacy of the pioneer drug).

247 *See* J. Jonas Anderson, *Secret Inventions*, 26 BERKELEY TECH. L.J. 917, 942–44 (2011) (citing Jeanne C. Fromer, *Patent Disclosure*, 94 IOWA L. REV. 539, 560–62 (2009), and Benjamin N. Roin, Note, *The Disclosure Function of the Patent System (or Lack Thereof)*, 118 HARV. L. REV. 2007, 2025 (2005)).

248 FDA does regulate diagnostic testing kits and companion diagnostics, but does not regulate testing services provided by individual laboratories. Steven Gutman, *The Role of Food and Drug Administration Regulation of in Vitro Diagnostic Devices—Applications to Genetics Testing*, 45 CLINICAL CHEMISTRY 746 (1999).

249 *See id.*

250 *See* Mayo Collaborative Servs. v. Prometheus Labs., Inc., 132 S. Ct. 1289, 1294 (2012).

251 Although black-box models may be static and stable, at least initially, complex big-data models in medicine may eventually be flexibly updated. This plasticity would make exclusivity mechanisms, whether regulatory or patent-based, less useful by further complicating the definition of a product.

3. Prizes

A third possibility to enhance innovation in algorithms is reliance on prizes or grants as a reward for innovation.[252] Grants and prizes each typically rely on the award of money—typically a fixed sum—to solve a defined problem. Under a grant regime, firms compete for monetary incentives that are then to be used to develop an innovation.[253] Under a prize regime, a monetary prize is offered to whichever firm can develop a solution to a defined problem.[254] Typically, the prize amount is fixed, though it need not be.[255] Such devices can avoid the requirement of exclusivity, either in situations where it is unavailable (e.g., when the innovation is unpatentable) or where free distribution is mandated as part of the incentive regime (e.g., where entering a prize competition or winning a grant requires relinquishing patent rights and committing to disclosure).

Prize and grant systems both require knowing the approximate contours of a defined problem with a defined solution and knowing the rough value of a solution to the problem.[256] Since personalized medicine in general and black-box medicine in particular are broad endeavors with significant implicit knowledge, clearly defining the problems and solutions appears particularly difficult. Goals could be defined very generally; for instance, any algorithm that decreases costs while

---

252 An extensive literature examines prizes and grants as alternatives to patents. *See* Abramowicz, *supra* note 181; Michael Kremer, *Patent Buyouts: A Mechanism for Encouraging Innovation*, 113 Q.J. ECON. 1137 (1998); Steven Shavell & Tanguy Van Ypersele, *Rewards Versus Intellectual Property Rights*, 44 J.L. & ECON. 525 (2001) (arguing for the superiority of an optional prize system); Joseph Stiglitz, *supra* note 181; Marlynn Wei, *Should Prizes Replace Patents? A Critique of the Medical Innovation Prize Act of 2005*, 13 B.U. J. SCI. & TECH. L. 25 (2007). *But see* F. Scott Kieff, *Property Rights and Property Rules for Commercializing Inventions*, 85 MINN. L. REV. 697 (2001) (arguing against prizes). For an overview of grants and prizes that places them in a taxonomy with patents and tax incentives, and argues that all four can set economic incentives that should be at base indistinguishable to rational firms, see Hemel & Ouellette, *supra* note 181, at 310–13. For an argument that patents and prizes, at least as applied, are largely indistinguishable, see Roin, *Intellectual Property Versus Prizes*, *supra* note 181. Roin also offers an extensive bibliography. *Id.* at 3–5. This literature has typically not included regulatory exclusivity among the menu of options, perhaps because its exclusivity model parallels that of patents; to the extent that regulatory exclusivity has benefits over patents for certain fields of technological innovation, it may obviate certain criticisms that lead at least some scholars to prefer prizes.

253 *See* Hemel & Ouellette, *supra* note 181, at 320–21.

254 *See id.* at 317–19.

255 For instance, instead of a fixed sum of money, a prize could be defined as a fraction of identifiable government savings attributable to the innovation. *Cf.* Nancy Gallini & Suzanne Scotchmer, *Intellectual Property: When Is it the Best Incentive System?*, 2 INNOVATION POL'Y & ECON. 51, 54 (2002) (arguing that prizes should be tied to social value); Earl L. Grinols & James W. Henderson, *Replace Pharmaceutical Patents Now*, 25 PHARMACOECONOMICS 355, 356 (2007) (proposing drug prizes tied to sales).

256 *See* Roin, *Intellectual Property Versus Prizes*, *supra* note 181, at 1026–29.

maintaining or increasing health measures. Such very broad (and very valuable) algorithms might be the most useful, but might also have the hardest time overcoming the competitive incentives of private parties to keep the algorithms secret. Goals could also be defined more narrowly; for instance, any algorithm that decreases the frequency of adverse reactions to taking a drug with a narrow therapeutic index by ten percent.

But the challenge of determining the optimal incentive size remains. The advantage of patents and other exclusivity regimes is that—at least ideally—the size of the reward should track the social value of the innovation.[257] Firms can use market information to project that value and invest accordingly. For prizes, whoever sets the prize—typically the government—usually must determine the eventual social value in advance; however, governments are typically not well-suited to this task.[258] Potentially, this problem could also be solved by basing the reward not on a specific dollar amount, but rather on a fraction of savings to government health programs like Medicare or Medicaid. This would increase proportionately with social value without requiring pre-estimation of the eventual size of the reward. However, many medically valuable uses are not particularly economical; for example, keeping a patient alive may lead to more costs in the future. Finally, prizes face considerable political economy problems; though many medical prizes have been proposed, implementation follows far behind.

Overall, although the specifics of implementation will require considerable care, prizes appear to be an attractive alternative to more traditional exclusivity incentives for the development of black-box medicine models. Achieving the right level of specificity and project definition is challenging, but that challenge also arises with patents and regulatory exclusivity regimes. Moreover, prizes can be precisely tailored and can be structured to require disclosure so as to enable continued cumulative innovation.

---

257 Dan L. Burk & Mark A. Lemley, *Policy Levers in Patent Law*, 89 VA. L. REV. 1575, 1580 (2003).

258 Grants face the same type of problem, though the grant-making organization must accurately estimate the cost of the innovation rather than the social value. Grants have other advantages: they are frequently used in biomedical research to incentivize innovation, and are, therefore, familiar; they leverage a social discount rate which is typically lower than private discount rates; and they avoid capital constraints and risk aversion. *See* Hemel & Ouellette, *supra* note 181, at 308. However, like prizes, they do not increase with the size of the eventual social welfare gain of the innovation, and therefore, face additional steps in guiding the allocation of innovative effort among projects.

C. *Incentives for Validation*

Finally, incentives are needed for model validation.[259] Unlike both traditional medical development and explicit personalized medicine, black-box medicine cannot readily be validated in standard, straightforward ways. However, black-box medicine still needs validation to ensure reliability. Instead of scientific understanding, clinical trials, and postmarket surveillance, the validation of complex, implicit black-box medicine models require validation through other computational mechanisms. Developing methods for that validation, and ensuring they are consistently applied, is an important piece of the innovation policy picture.

Innovation policy should ensure that appropriate incentives exist to drive validation. A bounty could be implemented for external validation (with standards likely set by FDA). Bounties could be set as a small fraction of the overall revenues of the model—as part of the initial regulatory exclusivity bargain, if one exists—paid by the original developer. The size of the reward would then roughly increase with the overall value of the model.[260] Rewards for confirmatory validation would ideally decrease asymptotically, so that initial validation would be much more valuable than further confirmation, but that any confirmation over a particular validity threshold received at least *some* reward. This could be set to ensure that the overall fraction of originator revenue that could be siphoned to incentivize validation would be capped.[261] On the contrary side, rewards for finding problems should also exist, and should likely not decrease with repetition.[262]

As an additional factor, concerns about validation are exacerbated when data and models are kept secret and proprietary. Implicit models are difficult to validate for the reasons described above, more difficult without access to the modeling code, and extremely difficult without access to the data on which the model was based. Thus, ensuring

---

259 *See supra* Section II.B.3.

260 One challenge is that focusing on monetary goals, whether revenue-based or savings-based, might focus incentives on models which deal primarily with costs rather than health improvements. If the principal goal of black-box medicine is cost-reduction, this focus would be unproblematic. However, if—as seems likely—improving health outcomes is either a primary objective of black-box medicine or at least an important ancillary objective, then an alternate path to valuing validation would be needed, or some combination of monetary savings and health outcomes.

261 For instance, for a validation cap of two percent, the first validator to pass a certain threshold could receive one percent, and then each subsequent validator could receive half the amount of the previous validator; the sum of these fractions converges to two percent.

262 The incentives available for challenges to models might be expected to decrease naturally; if a model is called into question, its value presumably decreases and any fixed fraction of that value would also decrease.

disclosure is important to enable not only development and cumulative innovation, but also validation of existing models.

Overall, the appropriate balance of innovation incentives for the development of black-box medicine requires significant and detailed further work. However, an optimal final landscape might include some push to assemble useful information, either via a public or public-private enterprise, tailored prizes to help drive algorithm development, and bounties for the purposes of third-party validation. In the latter two categories, the prosaic solution of increased grant funding for academic model development may also provide a significant boost in an area where the incentive needs are significant but not excessively large.

CONCLUSION

Overall, black-box medicine offers immense promise for changing the way medicine is practiced and the way medical technologies are created and deployed. However, the growth of black-box medicine requires an active and effective innovation policy. The current intellectual landscape in the United States creates problematic incentives that encourage firms to keep data secret and to focus on simple drug-device linkages, rather than developing the necessary capabilities to develop black-box medicine. This Article has suggested a few ways to change that innovation policy on the path to the major economic and health benefits of the next step in personalized medicine.

More generally, this Article stands along previous work to suggest that our broad-brush innovation system has problematic implications on the ground as it is applied to different questions of innovation in different industries.[263] The pharmaceutical and biomedical industries are typically characterized as areas where patents work fairly well; other industries, like software, are characterized as areas where patents work much less well to drive innovation.[264] This Article argues for greater

---

263 *See, e.g.*, ADAM B. JAFFE & JOSH LERNER, INNOVATION AND ITS DISCONTENTS: HOW OUR BROKEN PATENT SYSTEM IS ENDANGERING INNOVATION AND PROGRESS, AND WHAT TO DO ABOUT IT (2004) (defending industry-neutrality of patent laws); Dan L. Burk & Mark A. Lemley, *Is Patent Law Technology-Specific?*, 17 BERKELEY TECH. L.J. 1155 (2002); Michael W. Carroll, *One for All: The Problem of Uniformity Cost in Intellectual Property Law*, 55 AM. U. L. REV. 845 (2006); Michael W. Carroll, *One Size Does Not Fit All: A Framework for Tailoring Intellectual Property Rights*, 70 OHIO ST. L.J. 1361 (2009); William Fisher III, *The Disaggregation of Intellectual Property*, HARV. L. BULL., Summer 2004, at 24 (arguing for more industry-specific patent laws); Clarisa Long, *Our Uniform Patent System*, FED. LAW., Feb. 2008, at 44 (defending industry-neutrality of patent laws); Peter S. Menell, *A Method for Reforming the Patent System*, 13 MICH. TELECOMM. & TECH. L. REV. 487 (2007) (arguing for more industry-specific patent laws); Price, *Making Do in Making Drugs*, *supra* note 233.

264 *See, e.g.*, Edwin Mansfield, *Patents and Innovation: An Empirical Study*, 32 MGMT. SCI. 173 (1986) (reporting different rates of patent importance in different industries); Wesley M.

nuance and granularity even within industries. Drug manufacturing responds differently to patent and regulatory incentives than drug discovery and development,[265] development of new uses responds differently than developing initial uses,[266] and, as I have argued here, simple diagnostic tests respond differently to patent incentives than complex implicit algorithms. Setting incentives right, and directing innovation policy accordingly, is key to moving forward toward the future of medicine.

Cohen et al., *Protecting Their Intellectual Assets: Appropriability Conditions and Why U.S. Manufacturing Firms Patent (or Not)* (Nat'l Bureau of Econ. Research, Working Paper No. 7552, 2000), http://www.nber.org/papers/w7552 (same).

265 Price, *Making Do in Making Drugs*, *supra* note 233.

266 Roin, *Solving the Problem of New Uses*, *supra* note 13.