Risk Regulation and Innovation: The Case of Rights-Encumbered Biomedical Data Silos

Arti K. Rai
Duke Law School

Follow this and additional works at: http://scholarship.law.nd.edu/ndlr
Part of the Intellectual Property Law Commons

Recommended Citation
92 Notre Dame L. Rev 1645 (2017)
RISK REGULATION AND INNOVATION:  
THE CASE OF RIGHTS-ENCUMBERED  
BIOMEDICAL DATA SILOS

Arti K. Rai*

INTRODUCTION .................................................. 1642
I. BIOMEDICAL DATA: CHALLENGES OF CONCENTRATION AND  
   FRAGMENTATION ........................................... 1646
   A. Overlapping, Adjacent, and Fragmented Intellectual Property  
      Rights ................................................ 1646
   B. Diagnostic Data Silos: Beyond Intellectual Property Rights . 1651
II. AGGREGATING BIOPHARMACEUTICAL TRIAL DATA ............ 1655
   A. Biopharmaceutical Trial Data in the United States: Basic  
      History, Law, and Policy ............................... 1655
   B. Recent Developments at the FDA and EMA ................... 1658
III. DIAGNOSTIC TESTING DATA ............................... 1663
   A. The Basic Regulatory Scheme ............................ 1663
   B. Looking to the Future ................................. 1664
CONCLUSION .................................................... 1667

© 2017 Arti K. Rai. Individuals and nonprofit institutions may reproduce and  
distribute copies of this Article in any format at or below cost, for educational purposes, so  
long as each copy identifies the author, provides a citation to the Notre Dame Law Review,  
and includes this provision in the copyright notice.

* Elvin R. Latty Professor of Law, Duke Law School; Co-Director, Duke Law Center  
  for Innovation Policy. Thanks to Jorge Contreras, to participants in the Notre Dame Law Review  
symposium, and to participants in workshops at Fordham Law School and the University of Toronto Law School for helpful comments. From 2013–2015,  
I was a member of an Institute of Medicine (now “National Academy of Medicine”)  
committee that issued a report on clinical trial data sharing. See COMM. ON STRATEGIES FOR  
RESPONSIBLE SHARING OF CLINICAL TRIAL DATA, INST. OF MED., SHARING CLINICAL TRIAL  
DATA: MAXIMIZING BENEFITS, MINIMIZING RISK (2015) [hereinafter SHARING CLINICAL TRIAL  
DATA], http://nap.edu/18998. However, this Article is based on publicly available  
information only.
Introduction

Overlapping legal claims pervade data-based innovation. Because of its heavily regulated status, and its potential to improve human welfare, innovation based on biomedical data is a particularly notable example.

As in most other areas of innovation, intellectual property rights can overlap. Patents and trade secrecy can cover the same inventive territory simply because the U.S. Patent and Trademark Office (USPTO), litigants, and courts find it excessively expensive to police the patent statute’s complex disclosure requirements.

Moreover, precisely how much disclosure should be required is a policy question. Patent cases have sometimes allowed patent applicants to claim an invention that arises after patent filing without requiring disclosure of this after-arising invention. Meanwhile, the subsequent technology or information could itself be the subject of separate intellectual property, whether trade secrecy or an improvement patent.

While the caselaw trend has moved away from very broad rights, this movement has not precluded “adjacent” intellectual property rights. In these cases, the narrower patent claims do not formally cover the follow-on innovation. However, to the extent that even patent claims of moderate scope on a key research tool cannot be “invented around,” and the claims can be enforced, the practical effect may be similar.

Depending on the precise nature of the follow-on innovation generated, adjacent rights may include subsequent patents, trade secrecy, or regulatory exclusivities administered by the Food and Drug Administration (FDA). In the case of biomedical data—the focus of this Article—the adjacent right typically involves either trade secrecy or regulatory exclusivity. Both overlapping and adjacent rights have generated controversy regarding the possibility of excessive control by the initial rights holder.

In 2012 and 2013, when cases involving diagnostic testing patents ultimately brought these controversies to the U.S. Supreme Court, the Court not only struck down the individual patents involved in those cases but also suggested more generally that patents on diagnostic tools might confer undue control over future research. The Court used as its cudgel the argument that the territory covered by the patents failed to constitute patent-eligible subject matter under the so-called “law of nature” or “product of nature” doctrines.

While these Supreme Court cases address concern about excessive control by a single rights holder, there is an obvious flip side. Concentrating patent ownership in a single firm allows that firm to aggregate all of the “adjacent” data generated through use of the patent. With patents no longer

3 Ass’n for Molecular Pathology, 133 S. Ct. at 2116.
serving as “data aggregators”⁴ in the area of diagnostic testing, the problem of data fragmentation in this area may become even more acute.⁵ In diagnostic testing, as in other areas of biomedicine, large data sets promote cumulative innovation.⁶

Notably, once biomedical data becomes fragmented, overcoming fragmentation creates its own legal overlap challenge. Even with the recent Supreme Court cases, biomedical data pooling—like pooling of scientific data in other areas⁷—will face multiple intellectual property obstacles. These include both formal intellectual property—trade secrets and narrow patents—as well as reluctance on the part of academic researchers to share what they view as their competitive advantage.⁸

Biomedical data pooling also has to overcome relatively unique challenges posed by health privacy law. Privacy law—and related regimes of informed consent—produce an additional set of claimants, leading some to worry that these legal regimes will create a “tragedy of the anticommons” as formidable as that postulated by Michael Heller and Rebecca Eisenberg in their discussion of the challenge of assembling fragmented patent rights.⁹

Scholars have suggested both private- and public-sector responses to the pooling challenge. On the private-sector side, some have argued in favor of stronger individual rights. On this view, individual consumers accorded control or even property rights over “their” health data might be incentivized to


⁵ To be sure, as discussed in Section I.A, the legacy of certain broad patents, such as the patents at issue in Association of Molecular Pathologists v. Myriad Genetics, could act as a limited counterforce. But the scope of this counterforce is unclear. Additionally, data aggregation without some level of external access is problematic as a policy matter. See infra Section I.A.

⁶ See, e.g., Eichler et al., infra note 105 and accompanying text.


⁸ Scholars have long discussed, and documented empirically, strong resistance to data sharing among biomedical researchers, even on data associated with published research. See, e.g., Eric G. Campbell et al., Data Withholding in Academic Genetics: Evidence From a National Survey, 287 J. AM. MED. ASS’N 473, 473 (2002). In turn, public funders have used sticks and carrots to encourage data sharing by their grantees. See, e.g., Rebecca S. Eisenberg & Arti K. Rai, Harnessing and Sharing the Benefits of State-Sponsored Research: Intellectual Property Rights and Data Sharing in California’s Stem Cell Initiative, 21 BERKELEY TECH. L.J. 1187, 1196–99 (2006).

create pools. While proposals for some level of consumer control are interesting, according millions of individuals property rights in data is likely to exacerbate transaction cost problems, not resolve them.

Other scholars have highlighted the potential role of institutional private-sector players like insurance companies. The increasing interest shown by insurance firms in using claims data to determine the therapeutic value of medical interventions they cover is a positive development. However, it is not clear that insurance firms would necessarily be motivated to create pools across firm boundaries.

As for the public sector, many scholars (myself included) have discussed mechanisms by which public funders of data resource creation can exercise informal or formal regulatory power to promote data pooling. But public-sector funding is always a scarce resource.

One set of regulatory actors that have received less attention than they merit is risk regulators. In the data pooling context, risk regulators are part of the problem. But they are also part of the solution.

On the one hand, aggressive moves by risk regulators to control perceived risks to privacy may impair possibilities for pooling. More positively, for a significant set of biomedical data—all pre-approval biopharmaceutical data and a significant subset of diagnostic data—risk regulators like the FDA are the default data aggregators. They can use this status to promote innovation based on pooled data.

The recent history of the FDA’s role in addressing rights-encumbered biopharmaceutical clinical trial data silos highlights both the agency’s promise and its limitations as a promoter of data-based innovation infrastructure. This history also highlights the status of data as a global public good, with


12 See, e.g., Jorge L. Contreras, Leviathan in the Commons: Biomedical Data and the State, in GOVERNING MEDICAL KNOWLEDGE COMMONS (Brett M. Frischmann et al., eds. forthcoming 2017); Arti K. Rai & Rebecca S. Eisenberg, Bayh-Dole Reform and the Progress of Biomedicine, 66 LAW & CONTEMP. PROBS. 289 (2003) (addressing NIH legal requirements for the pooling of biomedical data created with federal funding).

13 For earlier articles that do focus on the FDA, see infra sources cited in note 104. Scholars have also talked about the role of the FDA in promoting drug safety through the creation of post-marketing data infrastructure. See, e.g., Barbara Evans, Congress’ New Infrastructure Model of Medical Privacy, 84 NOTRE DAME L. REV. 585 (2009). In the post-marketing context, however, the risk regulator does not necessarily have a significant amount of control over the data.
FDA action being supplemented, and arguably superseded, by the actions of its European counterpart, the European Medicines Agency (EMA).  

The data exclusivity administered by risk regulators in the United States and other advanced economies emerges as an attractive mechanism for balancing the interests of initial and subsequent data generators. Indeed, for purposes of promoting data pooling, the temporal overlap between patents and data exclusivity is a feature, not a bug. A similar regime of exclusivity for diagnostic data would be worth considering.

Data exclusivity does not address legal claims that emerge from the law of health privacy and informed consent. These regimes raise major transaction cost challenges, particularly to the extent data holders cannot guarantee that de-identified or anonymized data is impervious to re-identification. Unfortunately, U.S. risk regulators’ recent tool of choice for addressing potentially identifiable data—broad consent to any future research use—provides a suboptimal response. Broad consent is superior to any attempt to confer property rights on raw data inputs that will be provided by millions of different individuals. But it is inferior to a pragmatic approach that emphasizes sanctions for illegitimate re-identification and protecting those whose data is used from harm.

On balance, the contribution of risk regulators has been mixed. To some extent, risk regulators, particularly in Europe, have capitalized on their pivotal aggregation role to address the challenges posed by rights-encumbered data silos. However, even in the political environment of the Obama Administration—where the White House was a prominent advocate of open government data, and generally viewed regulatory activity as compatible with, or even a promoter of, innovation—U.S. risk regulators were cautious.

In elucidating the role of risk regulators, this Article contributes another chapter to the rich legal and economic literature discussing whether, and how, risk and social regulation can promote socially valuable innovation.  


16 Exec. Order No. 13,563 § 3, 3 C.F.R. pt. 100 (Jan. 18, 2011) (titled “Improving Regulation and Regulatory Review” and requiring that “[e]ach agency shall also seek to identify, as appropriate, means to achieve regulatory goals that are designed to promote innovation”).

As contrasted with much of this literature, however, the focus here is not on innovation induced by efforts to comply with regulation but instead on regulation as a force for creating a data infrastructure for future innovation.

Part I of this Article briefly summarizes the history of overlapping and adjacent intellectual property rights in biomedical innovation. It then discusses the manner in which the Supreme Court’s reaction to such rights concentration may exacerbate legally-encumbered diagnostic data silos. Part II outlines the basic history of biopharmaceutical trial data silos as well as the core legal and policy arguments in favor of increasing access to the aggregated data held by risk regulators. It then discusses recent developments, including the stance taken by European risk regulators. Part III discusses lessons from the biopharmaceutical trial data experience and how some of these lessons may play out in current debates over diagnostic testing silos and overlapping rights.

I. BIOMEDICAL DATA: CHALLENGES OF CONCENTRATION AND FRAGMENTATION

A. Overlapping, Adjacent, and Fragmented Intellectual Property Rights

Both inadvertently and by design, overlapping and adjacent intellectual property rights pervade biomedical innovation. As an initial matter, patents and trade secrecy can overlap. To be sure, the text of the patent statute—which requires that a patent document’s disclosure show the ordinary scientist or technologist working in the area how to make and use the claimed invention—appears to frown on direct overlap. So does the language of certain Supreme Court cases. As a practical matter, however, because of institutional limitations within the USPTO and other patent-granting institutions, patent disclosure is often inadequate. Additionally, because the invalidation of a bad patent represents a public good from which all competitors benefit, competitor attempts to mount challenges to patents are likely to be undersupplied. Undersupply of disclosure-based challenges may be particularly likely because proving inadequate disclosure can be highly fact-intensive and complex.

More fundamentally, patent law has an ambivalent relationship with full disclosure. For example, the law allows patentees to impose on those who wish to use the disclosure some burden of experimentation, so long as it is
not “undue.” Some level of less than comprehensive disclosure might be justified on the grounds that the law wants to encourage potential users to engage in ex ante licensing that gives them access to this additional disclosure rather than ex post settlement of infringement claims.

In addition, patent caselaw has sometimes indicated that claims can, consistent with the disclosure requirements of the patent statute, cover technology that did not exist when the claim was filed. One famous example, decided by the Federal Circuit in 1989, involved a patent that broadly claimed a new polymer that it termed crystalline polypropylene. At the time the patent application was filed, the term crystalline polypropylene was understood in light of the type of molecule the applicant had identified—a polymer of low viscosity and low molecular weight. In contrast, by the time the patent was asserted, the crystalline polypropylene art encompassed polymers of high viscosity and high molecular weight. The Federal Circuit held that because the adequacy of the patent’s disclosure was judged in relation to the state of the art at the time of patent filing, the patent was valid. Its scope was not, however, limited to the date of patent filing—thus, the patent claim encompassed the later-invented polymers.

The Federal Circuit’s treatment of crystalline polypropylene is consistent with an economic vision, famously advocated by Edmund Kitch, that patents are usefully analogized to mining prospects and should therefore claim

21 In re Wands, 858 F.2d 731, 736–42 (Fed. Cir. 1988). Scholarly commentators have also debated at length how important full disclosure is, or should be, to the patent system. See, e.g., Jeanne C. Fromer, Patent Disclosure, 94 IOWA L. REV 539 (2009); Lisa Larrimore Ouellette, Do Patents Disclose Useful Information?, 25 HARV. J.L. & TECH. 545 (2012). Indeed, a recent issue of the Vanderbilt Law Review focuses exclusively on this question. See Symposium, The Disclosure Function of the Patent System, 69 VAND. L. REV. 1455 (2016). In his contribution to this issue, Nicholson Price argues for an “economic enablement” standard that would go well beyond current doctrinal requirements. See W. Nicholson Price II, Expired Patents, Trade Secrets, and Stymied Competition, 92 NOTRE DAME L. REV. 1611 (2017). Price uses as an example of economic enablement the possibility that firms with diagnostic patents would be required, after the patent has expired, to disclose genetic variant data revealed through use of the patent. Id. at 1632–34, 1640.

22 Relatedly, to the extent that full disclosure includes transfer of know-how that is difficult to codify, ex ante patent licensing may provide an efficient mechanism for transferring this know-how. See ASHISH ARORA ET AL., MARKETS FOR TECHNOLOGY: THE ECONOMICS OF INNOVATION AND CORPORATE STRATEGY 116–18 (2001) (discussing patents as rights that a licensor can use as a “hostage” against misappropriation of know-how by the licensee).

23 See U.S. Steel Corp. v. Phillips Petroleum Co., 865 F.2d 1247 (Fed. Cir. 1989); see also In re Hogan, 559 F.2d 595 (C.C.P.A. 1977).

24 See U.S. Steel, 865 F.2d at 1250.

25 Id.

26 Id. at 1252.

27 Id. at 1250–52.

large areas of research territory yet to be mined. Kitch’s theory does not envision the possibility of improvement patents within the scope of the dominant patent. Thus, it would presumably create overlap only between patents and trade secrecy. As it happens, however, because black letter patent law allows improvement patents within the scope of the dominant patent, overlapping patents are equally possible. These overlapping patents may be held by the same, or different, owners.

In recent years, the Federal Circuit’s newfound emphasis on the so-called written description language found in section 112 of the patent statute has caused some reduction in complete overlap. According to the Federal Circuit, written description requires that the patent filer show possession of the technology claimed at the time the patent was filed. Possession is generally defined through structure, and an expansive claim construction that goes beyond the structures that existed at the time of filing can render the patent invalid. Thus, for example, some prominent Federal Circuit cases—including the court’s en banc 2010 decision in *Ariad Pharmaceuticals, Inc. v. Eli Lilly & Co.*—have wielded written description to strike down patent claims that purport to cover not only identified structures (e.g., a receptor) involved in a disease process but also as-yet undiscovered drugs that could work to modulate the receptor.

From the standpoint of those wary of overlapping rights, written description is only a partial solution, however. First, challenges based on written description can be difficult for courts to understand and expensive to mount. Second, in at least some cases, even a relatively narrow patent on a key research tool may not be susceptible to “inventing around” by a potential competitor. In that case, the patent on the tool may create “adjacent” rights on follow-on innovation that uses the tool.

29 For a recent analysis of Federal Circuit cases allowing applicants to claim technology that has not yet been invented, see for example Kevin Emerson Collins, *Enabling After-Arising Technology*, 34 J. Corp. L. 1083 (2009).


31 Certain cases tightening the enablement requirement may have also reduced overlap. See Jacob S. Sherkow, *Patent Law’s Reproducibility Paradox*, 66 Duke L.J. 845, 877–78 (2017) (arguing that the Federal Circuit has recently applied enablement in a manner that requires patentees to claim narrowly).

32 598 F.3d 1336 (Fed. Cir. 2010) (en banc); see also Abbvie Deutschland GMBH & Co. v. Janssen Biotech, Inc., 759 F.3d 1285 (Fed. Cir. 2014); Biogen Idec, Inc. v. GlaxoSmithKline LLC, 713 F.3d 1090 (Fed. Cir. 2013).


For example, to the extent it could be enforced against infringers, even a relatively narrow patent claim to an important cell receptor could create a zone of adjacent rights in the molecules that modulate that receptor. Indeed, the heated debate over patents on research tools and platform technologies—and the corresponding debate over the need for a general-purpose research exemption—was fueled in part by concern about adjacent rights.

As it happens, because many research tool patent owners were either unable or unwilling to enforce their rights, concerns about excessive control were mitigated. But when patentees did enforce their rights—as they sometimes did on diagnostic patents—controversy flared well beyond the relatively insular boundaries of the patent world.

Two cases, Mayo Collaborative Services v. Prometheus Laboratories, Inc. and Association for Molecular Pathology v. Myriad Genetics, Inc., brought these controversies to the attention of the Supreme Court. In both cases, the Court unanimously struck down patents that the challengers and their allies argued created chilling effects for follow-on research. In the Mayo case, Justice Breyer discussed at length undue preemption of follow-on research as a reason to invalidate the patents. Although Justice Breyer’s opinion did not specifically discuss adjacent rights over data, the preemptive effect would presumably encompass adjacent rights over data generated by the diagnostic test. The Court used as its guide to undue preemption whether the subject matter claimed constitutes a law or product of nature under section 101 of the patent statute.

Many commentators (myself included) have argued that the Court failed to articulate how section 101 overlaps with, or is distinct from, other inquiries within the patent statute that govern patent scope (and hence preemption) or, indeed, the ability to secure a patent in the first instance. As noted, however, unlike the fact-intensive and complex inquiries required by other
sections of the patent statute, a section 101 inquiry is considered a question of law and thus can be deployed early and inexpensively by the USPTO, by patent challengers, and by courts. Indeed, the empirical evidence thus far suggests significant use of section 101 to reject patents in areas related to diagnostics, even at the resource-constrained USPTO.43

Although the effects of Mayo and Myriad require further empirical study, plausible effects may include fewer patents and/or substantially narrower patent claims. Additionally, firms may further emphasize secrecy for diagnostic research tools. Both narrower claims and secrecy will presumably create more data fragmentation. Moreover, like all data, this fragmented diagnostic data will be subject to claims of trade secrecy protection under both longstanding state trade secret law and now the federal Defend Trade Secrets Act (DTSA).

In a few cases, the legacy effect of patents now deemed invalid may mitigate fragmentation. For example, the BRCA1 and BRCA2 diagnostic gene-testing patenting portfolio has given Myriad Genetics a database of test results from about 1.5 million patients.45 According to recent press accounts, the Myriad database has 17,000 BRCA variants in its proprietary database, and 46,000 gene variants in total.46 However, even in the case of Myriad, this legacy effect has not forestalled substantial competition.47 In any event, as many commentators have noted, data aggregation that maintains indefinite secrecy is far from ideal. Although entities that hold aggregated data can do research, the credibility of this research will be limited by the fact that they cannot release the data that undergirds the research.48


44 Defend Trade Secrets Act of 2016, Pub. L. No. 114-153, 130 Stat. 376 (to be codified at 18 U.S.C. §§ 1832–39, 1961). Without eliminating or preempting state trade secret rights, the DTSA makes the federal courts available for parties to bring civil actions for misappropriation of trade secrets. Under both the DTSA and the state law adopted by almost every state, the information in question must neither be known nor readily ascertainable by competitors, and it must derive economic value from such secrecy. The owner must also have taken reasonable measures to maintain secrecy. This definition is capacious enough to allow claims that siloed data qualifies as a trade secret. Trade secrets are protected against “misappropriation,” which is essentially a tort-type action wherein the accused party acquires the trade secret by improper means such as theft, bribery, misrepresentation, breach, or inducement of a breach of a duty to maintain secrecy.


46 Turna Ray, Genomic Variant Data Sharing Gains Support; Collaboration Seen as Key to Interpretation Challenge, Genome Web (May 2, 2016), https://www.genomeweb.com/informatics/genomic-variant-data-sharing-gains-support-collaboration-seen-key-interpretation.

47 See id.

48 Several recent articles address the question of whether, and how, law should address large legacy silos such as the Myriad silo. See, e.g., Robert Cook-Deegan et al., The Next Controversy in Genetic Testing: Clinical Data as Trade Secrets?, 21 Eur. J. Hum. Genetics 585.
But patents and trade secrecy are not the only legal regimes relevant to data. In addition, complex legal regimes governing health information privacy and informed consent create a plethora of legal hurdles. The next Section outlines these hurdles.

B. Diagnostic Data Silos: Beyond Intellectual Property Rights

In the context of human biomedical data, legal regimes conferring informed consent and privacy protections overlap with intellectual property claims, thereby further fortifying silos. Modern law governing human-subjects research originates in a number of troubling twentieth-century research scandals, including a notorious study (“the Tuskegee study”) in which researchers from the U.S. Public Health Service observed the progression of syphilis in poor African-American research participants without informing them of the fact they had the disease or of available treatments. Accordingly, both the Common Rule that governs federally-funded human-subjects research, and separate regulation for clinical trials subject to the jurisdiction of the FDA, place a high premium on informed consent.

Notably, human-subjects research regulations apply not only to interventional research involving interactions between a researcher and a living individual but also to secondary research on human data and biospecimens that have already been collected. Many commentators have noted that rules governing interventional research, which involves a risk of physical harm, do not necessarily function well when exported to the realm of purely informational research. Even so, the Common Rule continues to govern such research.

In the context of informational research, the relevant risk is breach of privacy, with attendant potential for harmful action by employers, insurers, law enforcement, and other institutions that receive access to the data. Following this logic, the Common Rule exempts informational research if the relevant information has been “recorded by the investigator in such a...”

---

49 See generally James H. Jones, Bad Blood (2d ed. 1993).
52 See 21 C.F.R. § 50.20 (FDA regulation); 45 C.F.R. § 46.116 (Common Rule).
53 See, e.g., Inst. of Med., Beyond the HIPAA Privacy Rule: Enhancing Privacy, Improving Health Through Research 7 (Sharyl J. Nass et al. eds., 2009) (“[i]n information-based research that relies solely on medical records and stored biospecimens, the research participant faces no risk of direct physical harm.”).
54 Existing law also places stricter restrictions on the secondary use of data for research than on its use in other contexts, such as public health practice. Stricter regulation of research use is grounded in the view that research, unlike other uses, aims for generalizable knowledge and does not purport directly to benefit research participants. See, e.g., Evans, supra note 13, at 615–19. Whether the law should draw sharp distinctions between different uses of data is an important question, but one that is beyond the scope of this Article.
manner that subjects cannot be identified, directly or through identifiers linked to the subjects.” 55 Similarly, the Privacy Rule promulgated pursuant to the Health Information Privacy and Accountability Act (HIPAA) of 1996 allows a “covered entity” (e.g., a provider or health insurance plan) that seeks to transfer data for research purposes to do so if the data has been subject to “de-identification.” 56

Because of the need for removal of identifiers, compliance with the Common Rule and the HIPAA Privacy Rule imposes a tax on sharing data. At least for certain types of data, this tax can be relatively modest, however. 57 Indeed, removal of identifying information, and contractual prohibitions against re-identification, have been the mechanism by which some prominent efforts at pooling clinical trial data have negotiated privacy and informed consent hurdles. For example, the Clinical Study Data Repository (CSDR), a private-sector platform that gives researchers access to patient-level clinical trial data from over a dozen pharmaceutical firms, is able to do so because it has “anonymized” the data, and requires researchers who access the data to refrain from re-identification. 58 Moreover, as discussed in Part II, the EMA is currently focusing on data sets that do not appear to create significant re-identification issues.

Going forward, the more significant challenge for pooling under the current Common Rule/HIPAA regime is the concern that certain types of important health information, particularly genomic information, are quite susceptible to re-identification. A number of scientific researchers have published articles demonstrating re-identification based on genomic data.59

Indeed, the Office of Human Research Protections (OHRP)—the component of the Department of Health and Human Services (HHS) responsible for administering the Common Rule—has proposed revising the rule to require written consent for secondary research on all biospecimens, even

55 45 C.F.R. § 46.101(b)(4).
56 HIPAA’s Privacy Rule has a two-pronged method for deeming information “de-identified.” Either a person “with appropriate knowledge of and experience with generally accepted statistical and scientific principles” must certify the information de-identified or a list of identifiers identified in the Rule must be removed. See 45 C.F.R. § 164.514(b) (2016).
57 For example, according to Johnson & Johnson, which is supporting the Yale Open Data Access project, the annual cost of maintaining a secure computer infrastructure is 200 to 330 times that of de-identifying data from a clinical trial. E-mail from Karla Childers, Senior Dir. of Strategic Projects, Johnson & Johnson, Inc., to Arti K. Rai (July 20, 2016) (on file with author). In cases where de-identification makes data less usable as a scientific matter, the tax is greater.
59 See, e.g., Melissa Gymrek et al., Identifying Personal Genomes by Surname Inference, 339 SCIENCE 321 (2013); Nils Homer et al., Resolving Individuals Contributing Trace Amounts of DNA to Highly Complex Mixtures Using High-Density SNP Genotyping Microarrays, 4 PLOS GENETICS e1000167, e1000167 (2008); see also Bohannon, Genealogy Databases Enable Naming of Anonymous DNA Donors, 339 SCIENCE 262, 262 (2013) (summarizing studies).
specimens that are considered nonidentifiable according to traditional standards.60 Instead of allowing for an exemption based on de-identification, the proposed rule would encourage researchers who collect biospecimens to seek broad consent for all future research uses.61

For its part, the National Institutes of Health (“NIH”) appears to have already decided that it does not consider any type of genomic data exempt from the consent requirement. NIH’s “Genomic Data Sharing Policy,” published on August 28, 2014,62 states that NIH “expects investigators generating genomic data to seek consent from participants for future research uses and the broadest possible sharing.”63 As justification for its position, NIH’s policy statement highlights the aforementioned research studies demonstrating re-identification.64

In theory, standardized, broad consent might appropriately respect privacy without imposing large transaction costs.65 However, as the NIH policy statement itself acknowledges, research participants may not necessarily agree to such broad consent, “potentially biasing studies if certain populations were less likely to consent to broad use of their data.”66 For this reason, NIH states that it will “continue to accept data from studies with consents that stipulate limitations on future uses and sharing.”67 But merging data gathered under multiple different consent terms is likely to pose significant transaction cost challenges.

Additionally, as discussed further below, both the NIH approach and the proposed Common Rule approach appear to give up too readily on imperfect de-identification and anonymization. They fail to recognize contexts in which data pooling is unlikely to result in re-identification or the alternative of concerted efforts to deter illegitimate re-identification. Indeed, a recent National Academy of Sciences (NAS) report was quite critical of the proposed Common Rule revision, suggesting instead a system of “explicit sanctions against investigators and institutions seeking to re-identify biospecimen

---

61 Id.
63 Id. at 51,347.
64 Id. (noting that the “evolution of genomic technology and analytical methods raises the risk of re-identification”). Notably, NIH practices with respect to biological specimens and genomic data have also been put under scrutiny by journalist accounts highlighting the absence of informed consent in the circumstances under which cancer cells were taken in 1951 from a biopsy of Henrietta Lacks’s tumor and transformed into the now famous research tools known as the HeLa cell lines. Ewen Callaway, Deal Done over HeLa Cell Line, 500 Nature 132 (2013).
65 Some empirical studies suggest that research participants value autonomy and consent highly. See S.B. Trinidad et al., Research Practice and Participant Preferences: The Growing Gulf, 331 Science 287 (2011). Thus, seeking broad consent is arguably consistent with participant preferences.
67 Id.
On the positive side, we also have a partial precedent—biopharmaceutical clinical trial data—in which risk regulators and data exclusivity administered by risk regulators figure prominently. As discussed in Part III, because the regulatory footprint in diagnostics is currently lighter than in therapeutics, the precedent is only partial. However, to the extent that safety concerns ultimately lead to a heavier regulatory footprint in diagnostics, biopharmaceutical trial data demonstrates the promise, and limits, of risk regulators as promoters of data pooling. Additionally, the mechanisms by which risk regulators have assisted biopharmaceutical data pooling are, on their own terms, important illustrations of the manner in which well-crafted regulatory regimes can promote innovation.

The next Part turns to biopharmaceutical trial data.

68 Comm. on Fed. Research Regulations & Reporting Requirements, Nat’l Acad. of Sci., Optimizing the Nation’s Investment in Academic Research 166 (2016) [hereinafter Nat’l Acad. of Sci., Optimizing], http://www.nap.edu/21824; see also Contreras, supra note 9. That said, from a transaction cost perspective, informed consent is much preferable to the alternative, intermittently advocated over the decades, of granting individual “sources” property rights over biospecimens or data inputs. See generally James Boyle, Shamans, Software, and Spleens (1996) (discussing the question of property rights for “sources” in contexts ranging from plant genetic resources to cell lines). For a recent proposal advocating property rights in medical record health data specifically, see for example Hall, supra note 10; Mark A. Hall & Kevin A. Schulman, Ownership of Medical Information, 301 J. Am. Med. Ass’n 1282, 1284 (2009).

Perhaps most notably, unlike property rights violations, allegations of a breach of informed consent do not yield remedies against third parties. Indeed, the famous California Supreme Court case of Moore v. Regents of the University of California, involving cell lines over which the plaintiff claimed a property interest, accepted Moore’s arguments regarding breach of informed consent but rejected his conversion claim in significant part because of concerns regarding third-party researchers. 793 P.2d 479, 480, 494 (Cal. 1990) (en banc). More generally, as many commentators have noted, because of regulators’ concerns about impeding downstream research, federal and state law generally refuses to recognize property rights in excised biological material and related data. See, e.g., R. Alta Charo, Body of Research—Ownership and Use of Human Tissue, 355 New Eng. J. Med. 1517, 1518 (2006); Rina Hakimian & David Korn, Ownership and Use of Tissue Specimens for Research, 292 J. Am. Med. Ass’n 2500, 2500 (2004). To be sure, as Jorge Contreras has recently emphasized, defendants might choose to settle lawsuits involving claims including breach of informed consent through actions that resemble property rule remedies. See Contreras, supra note 9, at 20–23. Thus, for example, in several lawsuits alleging state common-law breach of informed consent claims against researchers, defendants agreed to settlement terms that destroyed or returned the biological samples in their possession. However, a defendant’s choice to use remedies sounding in property rules (presumably because it would prefer not to pay large monetary damages) is not equivalent to court imposition of that remedy.
II. AGGREGATING BIOPHARMACEUTICAL TRIAL DATA

A. Biopharmaceutical Trial Data in the United States: Basic History, Law, and Policy

In order to enter the U.S. market, biopharmaceuticals must be approved by the FDA. In general, approval requires the submission of clinical trial data showing that the product in question is safe and effective. When generated by a private developer, biopharmaceutical trial data starts out as the developer’s trade secret. The question of how a submission to the FDA changes the data’s legal status is complex. Adding to the complexity, the legal regime for “large molecule” biologics data is different from, and less clear than, that for small molecule data. For clarity of exposition, I focus on small molecule drug data.

Historically, the FDA has taken the position that under both its organic statute, the Food, Drug, and Cosmetic Act (FDCA), as well as under the Trade Secrets Act (TSA) that applies to all federal agencies, it is prohibited from releasing drug trial data. Furthermore, in rulemaking done in 1974 to implement the Freedom of Information Act (FOIA), the agency concluded that it was not “feasible or practical to determine the differences, if any,” between what the agency is prohibited from disclosing under the FDCA and the TSA and what it is permitted to withhold from disclosure after a citizen request under FOIA. Thus, the FDA’s 1974 rule on what it is prohibited from disclosing largely mirrors the broad language of FOIA Exemption 4. The rule states that “[d]ata and information submitted or divulged to the Food and Drug Administration which fall within the definitions of a trade secret or confidential commercial or financial information are not available for public disclosure.”

With respect to the FDCA, which specifically bars the FDA from revealing only “any method or process which as a trade secret is entitled to protection,” the FDA’s position appears overly conservative. And even with respect to the TSA, which covers both “trade secrets” and “confidential statistical data,” this statute notes the caveat of disclosure being permissible if “authorized by law.” Other agencies have taken the caveat to mean that, in

---

71 See 39 Fed. Reg. 44,602, 44,634–35 (Dec. 24, 1974) (stating that the FDA’s position under the FDCA dates back to 1938 and that the Commissioner further concludes that the TSA precludes disclosure).
72 Id. at 44,612.
73 21 C.F.R. § 20.61(c) (2016) (formerly codified at 21 C.F.R. § 4.61(c)).
75 This point was made to the FDA in the 1974 FOIA rulemaking, but the agency rejected it on the grounds that it did not want to alter an interpretation it had maintained since 1938. 39 Fed. Reg. at 44,634–35.
76 Specifically, the TSA prohibits federal employees from releasing “to any extent not authorized by law” information they receive that relates to “trade secrets” or “confidential statistical data.” 18 U.S.C. § 1905 (2012).
those cases where their delegated authority gives them authority to promulgate rules that represent binding law, they may disclose information otherwise encompassed by the TSA by promulgating such rules. This interpretation of the TSA caveat was endorsed by the Supreme Court in its 1979 decision *Chrysler Corp. v. Brown*. This case also held that the TSA does not prohibit agencies from disclosing information that the agency may permissibly exempt under FOIA.

Whatever its legal merits, prior to the passage of the Hatch-Waxman Act of 1984 (“Hatch-Waxman”), the FDA’s conservative position may have been justified as a policy matter. As commentators have noted, molecules that might be promising drug candidates cannot always be reliably patented. Assuming no patents barred the action, competitors might have been able to use released data to immediately submit their own marketing applications (known as new drug applications, or NDAs).

The economic and legal calculus changed after the passage of Hatch-Waxman. Under the statute, an originator’s small molecule clinical trial data receives a five-year period of exclusive use post-marketing. Competitors must wait for that period to expire before they can market drugs that rely on the originators’ data as evidence of the safety and effectiveness of their own “bioequivalent” molecule.

Notably, in section 355(l) of Hatch-Waxman, Congress appears to have recognized this change of economic calculus. That section directs the FDA to make clinical trial data on approved NDAs available upon request to the public after the expiration of the exclusivity period “unless extraordinary circumstances are shown.” Hatch-Waxman also states that clinical trial data

---

78 *Id.* at 316–17.
80 See Rebecca S. Eisenberg, *Data Secrecy in the Age of Regulatory Exclusivity, in The Law and Theory of Trade Secrecy: A Handbook of Contemporary Research* 467, 469 (Rochelle C. Dreyfuss & Katherine J. Strandburg eds., 2011) (making this argument).
83 This five-year period applies both to drugs that are generic copies as well as drugs submitted under the so-called 505(b)(2) pathway. Drugs submitted under 505(b)(2) rely only in part on the innovator data. *Id.* § 105(a).
84 21 U.S.C. § 355(l)(1), (1)(E) (2012) (“Safety and effectiveness data and information which has been submitted in an [NDA] for a drug and which has not previously been disclosed to the public shall be made available to the public, upon request, unless extraordinary circumstances are shown . . . upon the effective date of the approval of the first [abbreviated new drug application (ANDA)] which refers to such drug or upon the date upon which the approval of an [ANDA] which refers to such drug could be made effective if such an application had been submitted.”).
on drugs whose NDA has been irrevocably rejected should be made available
absent “extraordinary circumstances.”\footnote{Id. § 355(l)(1)(A)–(C). For its part, the FDA has promulgated regulations that mir-
ror the language of the statute. \textit{See}, e.g., 21 C.F.R. § 314.108 (1999) (discussing exclusivity in
regulations promulgated by the FDA).}

More generally, even absent an affirmative request by a member of the
public, data exclusivity providing protection that survives disclosure would
appear to obviate the need to maintain secrecy over the data. Additionally,
this type of data exclusivity now exists in most major jurisdictions that require
clinical data for market entry.\footnote{See generally Jerome H. Reichman, \textit{Rethinking the Role
As in the United States, it attaches upon marketing and runs concurrently with any patent protection.

Those who resist data release have argued that such release might none-
theless damage the originator firm’s competitive prospects in countries that
lack regulatory protection for data.\footnote{\textit{Sharing Clinical Trial Data}, \textit{supra} note *, at 260.}
On this argument, competitors could take the data package released and use it immediately to compete with the
originator in those countries.\footnote{Id. at 261.} However, as documented in a 2015 Institute
of Medicine (IOM) report discussing clinical trial data sharing,\footnote{Id.} many of
these countries require little or no data submission in any event, instead relying
largely on the approval of the drug by the FDA or EMA. In these coun-
tries, the barrier (if any) to generic competition is patent protection secured
by the originator firm, not trial data.\footnote{These countries, which are found largely in the developing world, under-
standably make tradeoffs between access and innovation incentives different from those made in
richer countries.} In any event, as prescribed by section 355(l) of Hatch-Waxman, the regulator might release data after the exclusiv-
ity period in its jurisdiction has expired, thereby providing more secure pro-
tection to the initial data generator.\footnote{Pub. L. No. 98-417, § 102, 98 Stat. 1585, 1592 (1984). As discussed below, the EMA
policy does not delay release. Instead, it requires those who access the data to sign a con-
tract stating that they will not use the data for purposes of pursuing marketing
applications.}

Even after Hatch-Waxman, the FDA resisted efforts at data release. In a
1999 case where Public Citizen Health Research Group invoked section 355(l) to argue for data release, the agency took the dubious position that
the statutory language allowed it to withhold the same information that it was

Biopharmaceutical firms have also argued that commercially sensitive information
other than safety and effectiveness data might be included in the clinical study reports
(CSRs) submitted to regulators. This information could include company strategies for
dealing with regulators, strategies for conducting future clinical trials, and potential
exploratory endpoints. This information should, however, be susceptible to segregation
and redaction. The EMA policy, discussed below, operates accordingly.
permitted to withhold under Exemption 4 of FOIA.\textsuperscript{92} The FDA's stance appears to have chilled subsequent attempts to invoke section 355(l).\textsuperscript{93}

\textbf{B. Recent Developments at the FDA and EMA}

In the United States, Congress ultimately acted to mandate regulatory action on data release. Activity and discussion began in earnest after a very prominent scandal surrounding Merck’s suppression of adverse event data with respect to the drug Vioxx.\textsuperscript{94} Indeed, “[a]ccording to the bill’s sponsors, [the] drug provisions” of the 2007 FDA Amendments Act (FDAAA) “were a direct reaction to the Vioxx scandal.”\textsuperscript{95} The Food and Drug Administration Modernization Act of 1997\textsuperscript{96} had previously mandated registration in ClinicalTrials.gov of federally or privately-funded clinical trials.\textsuperscript{97} FDAAA went a significant step further, requiring reporting of summary results from non-phase 1 trials of FDA-approved drugs, biologics and devices, whether or not the results had been published.\textsuperscript{98} These summary results include demographic and baseline characteristics of the patient sample as well as primary and secondary outcomes.

Summary results are useful, even more so if (as has recently occurred) requirements to release summary results encompass trial results from drugs that are either not the subject of an NDA or are ultimately not approved.\textsuperscript{99} But the goals of aggregation are better served through access to more detailed information. One level of greater detail is the clinical study report (CSR), which is submitted to regulators as part of an NDA.\textsuperscript{100} CSRs, which

\begin{itemize}
  \item \textsuperscript{92} Pub. Citizen Health Research Grp. v. Food & Drug Admin., 185 F.3d 898, 901 (D.C. Cir. 1999). The district court in that case rejected the FDA’s position, arguing that it was contrary to the statutory language under step one of \textit{Chevron}. \textit{Id.} at 900. The appellate court determined it did not have to decide the question because section 355(l) did not cover the type of data (investigational new drug application data rather than data on an approved or rejected NDA) sought by Public Citizen. \textit{Id.} at 903.
  \item \textsuperscript{93} A search of Westlaw for cases discussing section 355(l) yielded no results after 1999.
  \item \textsuperscript{95} \textit{Id.} at 616.
  \item \textsuperscript{96} Pub. L. No. 105-115, 111 Stat. 2296.
  \item \textsuperscript{97} Pub. L. No. 110-85, § 801, 121 Stat. 823, 904–22 (2007). Registration requirements include a brief summary; intended for the lay public; a discussion of the trial’s primary purpose and design; the intervention name and intervention type; target number of subjects; and outcomes, including primary and secondary outcome measures. \textit{Id.} § 801(j)(2)(A)(ii), 121 Stat. 823, 906.
  \item \textsuperscript{99} The FDAAA authorizes HHS to promulgate regulations that expand the set of results that must be reported to include failed trials, and HHS has recently promulgated final regulations along these lines. Clinical Trials Registration and Results Information Submission, 81 Fed. Reg. 64,982, 64,982 (Sept. 21, 2016).
  \item \textsuperscript{100} See \textit{International Conference on Harmonisation; Guideline on Structure and Content of Clinical Study Reports}; Availability, 61 Fed. Reg. 37,320 (July 17, 1996).
\end{itemize}
are standardized across major regulatory jurisdictions, often run into the hundreds of pages, providing detailed descriptions of the trial, clinical end points, methods, and key data.\textsuperscript{101} Data at the level of the individual patient goes one step further, potentially revealing flaws in the CSR.\textsuperscript{102} 

While proponents of clinical trial data release initially focused on a very specific type of innovation—policing erroneous or even fraudulent claims about safety and efficacy\textsuperscript{103}—they subsequently came to realize that techniques for analyzing data had improved to the point where aggregated data could create a more general-purpose innovation infrastructure.\textsuperscript{104} Specifically, release of large amounts of data—including data on biopharmaceuticals that have failed either in clinical trials or at the FDA approval stage—could be invaluable for improving future clinical trial design, avoiding duplicative clinical trials on biopharmaceuticals, conducting subgroup analyses that highlight new research directions (e.g., the possibility that an approved or failed drug might have a new use in a particular population), and conducting comparative effectiveness analysis.\textsuperscript{105} 

In 2009, when the Obama Administration came into office, the White House added to pressure on the FDA when it promulgated with some fanfare an “Open Government” Directive. The FDA responded by setting up a Transparency Task Force.\textsuperscript{106} On the question of clinical trial data, this Task Force proposed to

\begin{quote}
convene a group of internal and external stakeholders to discuss the possible uses of non-summary safety and effectiveness data from product applications, the circumstances under which it would be appropriate for sponsors to disclose non-summary safety and effectiveness data from applications submitted to FDA, and if appropriate, the format and the method by which disclosure should occur.\textsuperscript{107}
\end{quote}

Whether this group was ever convened is unclear.

\textsuperscript{101} Id.
Meanwhile in Europe, the controversy over clinical trial data release was becoming quite heated. In June 2007, researchers from Norway associated with the Cochrane group—a collaboration of medical researchers based largely outside the United States who do not take money from pharmaceutical firms—had requested from the EMA CSRs for fifteen placebo-controlled trials on two anti-obesity drugs. Their goal was “to study selective publication by comparing protocols and unpublished results with those in published reports.” Over the next three years, the EMA resisted the researchers’ requests by citing, inter alia, protection of commercially confidential information (CCI) and the administrative burden involved. The researchers responded by filing a complaint with the European Ombudsman, who found that the reports did not contain CCI, and issued a June 2010 statement castigating the EMA. In August 2010, the EMA reversed position, stating that it would grant access to the reports.

Three months later, in November 2010, the EMA also announced a more general policy stating (in line with the Ombudsman’s recommendation) that it would make CSR data on drugs for which a final decision has been made available on request. Between 2010 and 2013, the EMA received 750 requests for document access, and it granted access to over two million pages of documents. Having decided to release data, the EMA then stood by its decision in the face of lawsuits by originator firms.

The initial EMA decision catalyzed a series of events that created further momentum in favor of data release. As a consequence of the new EMA policy, another group of Cochrane researchers were able to obtain CSRs for the prominent antiviral drug Tamiflu from the EMA. Based on these CSRs, which amounted to tens of thousands of pages, the researchers published a series of papers asserting that much of the information about Tamiflu contained in previously published papers was incorrect or misleading.

In a 2012 Policy Forum published in the Public Library of Science: Medicine, these Cochrane researchers summarized their findings and placed them in the context of the larger debate by noting that further study of unpublished clinical trial data had also “radically changed public knowledge of safety and loads/AboutFDA/Transparency/PublicDisclosure/GlossaryofAcronymsandAbbreviations/UCM212110.pdf.

109 Id.
110 Id. at d2687.
111 Id.
112 Id.
113 Id. at d2689.
114 Requests for clinical trial data in particular were most likely to come from the pharmaceutical industry, academics, and law firms. See Sergio Bonini et al., Transparency and the European Medicines Agency—Sharing of Clinical Trial Data, 371 New Engl. J. Med. 2452, 2452 (2014).
efficacy” for such important drugs as Avandia, Neurontin, and Vioxx.\textsuperscript{116} Notably, the EMA responded by agreeing.\textsuperscript{117} In that April 2012 document, the EMA also indicated that it was moving towards proactive disclosure.\textsuperscript{118}

In June 2013, the EMA released for public consultation a draft policy on publication and access to CSR documents. The most controversial aspect of the policy has been what information the EMA will consider CCI, which it allows to be redacted. Since October 2014, when the EMA released the final policy,\textsuperscript{119} it has also issued a number of detailed guidance documents on the question.\textsuperscript{120}

In general, the EMA has taken the position that most clinical trial data is not CCI and that requests for redaction should therefore be limited and should be justified through specific argument. The EMA’s approach appears measured and deliberate. For example, the EMA has quite appropriately specified that information regarding exploratory or secondary endpoints not specifically targeted in the trial (and thus probably not already patented) can be redacted.\textsuperscript{121} Allowing such redaction preserves the possibility that these subsequent uses could be patented and then tested more systematically in the future.\textsuperscript{122}

\textsuperscript{116} Id. at 2.

\textsuperscript{117} See Hans-Georg Eichler et al., Open Clinical Trial Data for All? A View From Regulators, 9 PLOS MED. 1, 1 (2012).

\textsuperscript{118} See id.


\textsuperscript{121} See October 2014 Policy, supra note 119, at 17–20.

\textsuperscript{122} In some cases, courts have held that the mere mention of a potential secondary use in an article is sufficient to render the use unpatentable. See, e.g., MEHL/Biophile Int’l Corp. v. Milgaum, 192 F.3d 1362 (Fed. Cir. 1999). The EMA discussion regarding redaction of exploratory objectives specifically recognizes this possibility. See October 2014 Policy, supra note 119, at 19 (“[E]xploratory objectives may include biomarkers that could be used as ‘hypothesis generating’ for future studies. At that stage there would not be enough information to file patent applications on these objectives until some data are available from clinical and non-clinical studies. Disclosing these exploratory objectives may preclude obtaining patents . . . .”). Nicholson Price and Timo Minssen note these protections but argue that the EMA policy may nonetheless render public and thus unpatentable “secondary effects of unrecognized importance.” W. Nicholson Price & Timo Minssen, Will Clinical Trial Data Disclosure Reduce Incentives to Develop New Uses of Drugs?, 33 Nature Biotech. 685, 686 (2015). While it is possible that CSRs might contain stray observations about secondary effects that go unrecognized even by the pharmaceutical firm that is redacting the records, such failure to recognize potentially valuable information is not likely to be pervasive.
Under the policy, after the EMA has approved a drug for which a marketing access application was received after January 1, 2015, associated CSRs will be made available. The first CSRs released under this policy were made available in late 2016.

Although the CSR data is being made available immediately upon drug approval, jurisdictions with data exclusivity regimes will presumably not allow it to be used to seek marketing approval until after the exclusivity period has passed. As further protection for originators, the EMA is watermarking the data and also requiring those who download data to sign a contractual agreement forbidding use of the data for marketing approval. Of course, an EMA policy that delayed data release until after a certain period of time had elapsed post-approval would have further protected originator incentives. As the EMA monitors the effects of policy implementation, it should consider whether adjustments to the data release timeline are appropriate.

The FDA has not matched the EMA’s flurry of activity. However, absent evidence (which has yet to emerge) that drug makers are avoiding seeking regulatory approval in Europe based on concerns about data release, the FDA may not need to. The EMA’s unilateral action creates a global public good that does not have to be duplicated by the FDA. More generally, the EMA’s action highlights the extent to which, in certain contexts, unilateral action by a single large regulator can set a global baseline.

The interest demonstrated by the EMA and FDA also appears to have spurred private-sector collective action in the shadow of regulatory threat. One of the most notable efforts is the Clinical Data Study Request (CDSR) project, which makes anonymized individual patient data from thirteen originator firms available to researchers. The project has its roots in an announcement by GlaxoSmithKline in October 2012, in which the firm stated its goal of seeing “a broad system develop where the clinical research community can access data from trials conducted by different organizations.”

To be sure, the EMA has not yet tackled individual patient data, which poses the biggest challenges with respect to privacy and consent. That said, the EMA decision to take on the data-release challenge in a stepwise fashion, with the first step addressing aggregated data that presents fewer privacy concerns, itself represents a lesson of sorts. The EMA approach contrasts with the sweeping decision taken by NIH to deem all genomic information identi-

---

123 See October 2014 Policy, supra note 119, at 6–8.

124 In 2013, the FDA did issue a request for comments on a proposal to release drug trial data in “masked” form such that the sponsor of the drug(s) in question could not be identified. Although the FDA ultimately did not go forward on this proposal, it has recently published some internal analyses that assess classes of drugs and also shed some light on its approval process. See, e.g., Steven N. Goodman & Rita F. Redberg, Opening the FDA Black Box, 311 J. Am. Med. Ass’n 361 (2014).

fiable, and require researchers who collect this information to obtain broad consent for any secondary use.

The next Part discusses diagnostic testing data and lessons from biopharmaceutical trial data that could be applied in that area.

III. Diagnostic Testing Data

Because the regulatory scheme governing diagnostics differs from that governing therapeutics, this Part begins with a discussion of the basic regulatory scheme for diagnostics.

A. The Basic Regulatory Scheme

Since 1976, when it was given jurisdiction over medical devices, the FDA has regulated as medical devices all diagnostic tests that are marketed as kits.\textsuperscript{126} Thus, the FDA has test data on marketed kits. In contrast, because the FDA has exercised “enforcement discretion” over so-called laboratory-developed diagnostic tests, or LDTs, it has LDT data only in those cases where the firms or laboratories performing the LDT voluntarily seek regulatory approval.\textsuperscript{127}

In recent years, the FDA has become quite concerned about the quality of LDTs it does not regulate, particularly complex LDTs that rely on non-transparent algorithms that analyze large amounts of genomic, proteomic, and other biomarker data. During the Obama Administration, it proposed LDT regulation, emphasizing that the current regulatory scheme, under which the Center for Medicare and Medicaid Services assesses a test’s analytical validity under the Clinical Laboratory Improvement Amendments (CLIA), does not assess the accuracy with which a test measures the presence or absence of a particular clinical condition.\textsuperscript{128}

More specifically, the FDA proposed a “risk-based” approach to the regulation of laboratory-developed tests.\textsuperscript{129} Under this approach, “low-risk” LDTs, LDTs for rare diseases and unmet needs, and “traditional LDTs” used and interpreted directly by physicians and pathologists working within a single institution responsible for patient care would be subject only to notification and/or registration and listing, and adverse event reporting.\textsuperscript{130} Premarket approval (PMA) requirements would not apply. For tests that were found moderate- or high-risk, applicable requirements would depend on whether the tests in question were considered to be the highest-risk (Class


\textsuperscript{128} Id. at 59,776, 59,779.

\textsuperscript{129} See id.

\textsuperscript{130} Id. at 59,777.
devices or more moderate-risk (Class II).\textsuperscript{131} According to the FDA, it was particularly concerned about LDTs in three categories: those that use complex, nontransparent algorithms; those that act like companion diagnostics—that is, tests that claim to guide selection of therapy; and those that have the same intended use as an FDA-approved medical device.\textsuperscript{132}

The FDA’s work in this area followed a prominent IOM report, issued in 2012, that recommended that the FDA “develop and finalize a risk-based guidance or regulation on bringing omics-based tests to FDA for review.”\textsuperscript{133} The IOM report also called for greater transparency in the data, metadata, analysis plans, computer code, and computational procedures at issue in complex genomics tests.

The FDA proposal for LDT regulation did not suggest, however, that the agency would move to promote transparency of the underlying data it receives from test developers. Instead, the FDA would provide a summary review of its approval determinations, as it does for all medical devices.\textsuperscript{134} As medical devices, LDTs would also be subject to the various FDAAA requirements outlined in Part II.

\textit{B. Looking to the Future}

The issue of data aggregation and transparency may come to a head in the context of diagnostic tests, including LDTs, that produce data on genetic variants. For such tests, aggregation of variant data is essential for achieving high-quality test interpretation. Indeed, a July 2016 draft guidance from the FDA specifically argues for development of “publicly accessible” genetic variant databases.\textsuperscript{135} As the draft guidance emphasizes, the diagnostic utility of a specific class of LDTs—cutting-edge, next generation sequencing (NGS) techniques—depends on the robustness of “genetic variant data aggregation.”\textsuperscript{136} The FDA states it will recognize these publicly accessible databases “as sources of valid scientific evidence to support the clinical validity of geno-

\textsuperscript{131} Id.

\textsuperscript{132} Id.

\textsuperscript{133} \textit{Comm. on the Review of Omics-Based Tests for Predicting Patient Outcomes in Clinical Trials, Inst. of Med., Evolution of Translational Omics: Lessons Learned and the Path Forward} 136 (Christine M. Michiel et al., eds. 2012) (emphasis omitted), http://www.nap.edu/13297.

\textsuperscript{134} Framework for Regulatory Oversight of Laboratory Developed Tests; Draft Guidance for Industry, Food and Drug Administration Staff, and Clinical Laboratories; Availability, 79 Fed. Reg. at 51,352; \textit{see also} \textit{Bo. on Health Sci. Policy, Inst. of Med., Genome-Based Diagnostics: Clarifying Pathways to Clinical Use} 8 (2012), http://www.nap.edu/13359.


\textsuperscript{136} Id. at 2.
type-phenotype relationships in FDA’s regulatory review of NGS-based
tests.”

Assuming that it is finalized in the future, the FDA’s draft guidance may
provide some impetus for test developers to contribute their own variant data
to publicly accessible databases. However, nothing in the FDA guidance sug-
gests that the agency will require test developers to contribute their data.
Test developers could presumably free ride on public data for purposes of
making arguments to the FDA, while keeping their own data private. Indeed,
commentators’ reaction to the draft guidance has focused on the dearth of
databases that currently meet the FDA standard. While NIH funding has
generated some public databases, most notably the ClinVar resource, more
funding will be needed to increase the number of variants of ClinVar, and for
further curation.

Moreover, as discussed earlier, NIH funding requirements now require
all researchers who generate genomic data to have obtained broad consent
ex ante to share any data, no matter now unlikely it is to be re-identified.
As a consequence, prospects for rapid growth of ClinVar and other NIH-
funded databases through contributions by federally-funded researchers are
unclear. More generally, proposed actions by the HHS Office of Human
Research to revise the Common Rule that governs all federally-funded
human-subjects research so as to require written consent for secondary
research on biospecimens will create challenges for data pooling. These
include, for example, pooling efforts that may arise as a consequence of the
much-touted Precision Medicine Initiative (PMI)—an effort to track a vast
variety of health information, including genomic data, on one million partici-
pants. Notably, although the PMI was begun under the Obama Adminis-
tration, Congress recently passed legislation, the 21st Century Cures Act,
which provides robust funding well into the next administration.

Perhaps not surprisingly, distinguished commentators like the NAS have
been quite critical of approaches that treat all research as human-subjects
research requiring ex ante consent. A recent NAS report notes that the “sys-
tem may, for example, be better served by explicit sanctions against investiga-
tors and institutions seeking to re-identify biospecimen sources by any
method, including linkage of genomic sequence data to identifiers, rather

137 Id. at 1.
138 Turna Ray, FDA Plans to Use Public Genetic Variant Databases for NGS Test Regulation
But Not Many May Qualify, GENOMEWEB (July 12, 2016), https://www.genomeweb.com/
sequencing/fda-plans-use-public-genetic-variant-databases-ngs-test-regulation-not-many-
may-qualify.
139 See id.
140 See supra notes 62–68 and accompanying text.
141 See 45 C.F.R. § 46.117 (2009).
142 See About the Precision Medicine Initiative, NAT’L INSTS. OF HEALTH, https://www.nih
143 Turna Ray, Cures Act Gives Precision Medicine Proponents, Diagnostics Companies Reasons
to Celebrate, GENOMEWEB (Dec. 29, 2016), https://www.genomeweb.com/molecular-diag-
than by redefining all research with de-identified biospecimens as human subject research." 144 Ideally, HHS and its component institutions would have more carefully analyzed the question of how to generate publicly accessible databases. 145

In fairness to the FDA, current law does not provide any regulatory data exclusivity for FDA-approved diagnostic tests, or for medical devices more generally. Thus, the policy environment differs from drug trial data disclosure. But the precedent of biopharmaceutical clinical trial data, particularly in Europe, suggests that a properly balanced approach to transparency of FDA-regulated medical device data could involve a framework similar to that provided by Hatch-Waxman. For a specified period of time after the device’s approval, data would be protected by exclusivity and perhaps even protected from release. After this time period, however, it would be available for aggregation, with all the attendant benefits for patient health and future innovation.

Because this properly balanced framework would require congressional action, some might argue that it is a nonstarter. But the 2017 change in administration has already put congressional action on the agenda. With the change in administration, the FDA has indicated it will table its efforts to finalize its LDT guidance. 146 Many have argued that the FDA’s organic statute does not currently provide adequate regulatory authority, and have threatened to sue if the FDA begins to enforce its proposal. 147 Going forward, therefore, Congress is the most likely possible venue for comprehensive regulatory action on all diagnostic tests.

If and when Congress prescribes some type of regulatory approval requirement for diagnostic tests that are not currently regulated, it should strongly consider mechanisms by which the regulators’ role as a data aggregator could be exploited for future innovation. To some extent, Congress has already done that in the biopharmaceutical trial data context. Moreover, as the example of the EMA in the biopharmaceutical trial data context demonstrates, action by just one large regulator suffices to create a global baseline of data access.

144 Nat’l Acad. of Sci., Optimizing, supra note 68, at 166; see also Contreras, supra note 9.
145 For a wide-ranging argument that component agencies of HHS like the FDA and NIH should coordinate to promote better health innovation policy, see Rachel Sachs, Administering Health Innovation Policy (unpublished manuscript) (on file with author).
CONCLUSION

The Supreme Court’s intervention in the Mayo and Myriad cases addressed one type of legal overlap—adjacent and therefore arguably overly robust rights held by a single entity. But the diagnostic data fragmentation that is exacerbated by these cases poses its own legal overlap challenge. For each diagnostic data silo, multiple legal claims and claimants must be addressed to achieve the benefits of pooling.

Commentators who have discussed the data aggregation challenge have generally focused on possibilities created through public funding, through collective action by research participants, or through pressure by payers. This Article has emphasized the important role of risk regulators, most notably the partial precedent offered by biopharmaceutical clinical trial data. It has also highlighted the challenges created by U.S. regulators’ decision to require consent even for secondary research on data.

While U.S. risk regulators may have missed the mark on consent, the FDA has taken some steps forward on biopharmaceutical data pooling. Moreover, because data is a global public good, the actions of the EMA create a global baseline for access. The private-sector data pooling spurred by threats of regulatory action is also a positive development. Finally, in those advanced economies that have such exclusivity, data exclusivity that runs concurrently with any patents emerges as an attractive mechanism for balancing the interests of initial and subsequent data generators.