Governing secondary research use of health data and specimens: the inequitable distribution of regulatory burden between federally funded and industry research

Kayte Spector-Bagdady*

Corresponding author. E-mail: kaytesb@med.umich.edu

ABSTRACT

Some of the most promising recent advances in health research offer opportunities to improve diagnosis and therapy for millions of patients. They also require access to massive collections of health data and specimens. This need has generated an aggressive and lucrative push toward amassing troves of human data and biospecimens within academia and private industry. But the differences between the strict regulations that govern federally funded researchers in academic medical centers (AMCs) versus those that apply to the collection of health data and specimens by industry can entrench disparities. This article will discuss the value of secondary research with data and specimens and analyze why AMCs have been put at a disadvantage as compared to industry in amassing the large datasets that enable this work. It will explore the limitations of this current governance structure and propose that, moving forward, AMCs should set their own standards for commercialization of the data and specimens they generate in-house, the ability of their researchers to use industry data for their own work,

* Assistant professor, Obstetrics and Gynecology, University of Michigan Medical School; associate director, Center for Bioethics & Social Sciences in Medicine. This work was supported by the National Human Genome Research Institute (K01HG010496), the National Cancer Institute (R01CA21482904), and the National Center for Advancing Translational Sciences (UL1TR002240). Thank you to the participants in the Saint Louis University School of Law Center for Health Law Studies and the American Society of Law, Medicine and Ethics’ 2019 Health Law Scholars Workshop, as well as Professors Sharona Hoffman, Paul Lombardo, Kirsten Osterr, Elizabeth Pendo, W. Nicholson Price II, Tara Sklar, and Raquiah Yearby for their insightful comments on a previous draft. All errors are my own.
and baseline informed consent standards for their own patients in order to ensure future data accessibility.

**KEYWORDS:** Research, informed consent, Common Rule, data, HIPAA

## I. INTRODUCTION

Some of the most promising recent advances in health research—including 'precision medicine' and other genetic, machine learning, and big-data protocols—offer opportunities to improve diagnosis and therapy for millions of patients.\(^1\)\(^2\) They also require access to massive collections of health data and specimens to analyze the correlations between genetic variants, behaviors, environment, and health outcomes.\(^3\) This need has generated an aggressive and lucrative push toward amassing troves of human biospecimens, health, and health-proxy data.\(^4\) But there are profound differences between the strict federal human subjects research regulations that govern federally funded researchers versus the rules that apply to the collection of health data and specimens by industry. This difference can entrench disparities in the value and breadth of data available as between the two. But why would the US government make it harder for federally funded researchers to accomplish their work than industry-funded ones?

In 2019 the US Department of Health and Human Services substantively revised the ‘Common Rule’ portion of the Human Subjects Research Regulations (which generally regulate federally-funded researchers) for the first time since its original conceptualization in the 1980s.\(^5\)\(^6\) The revisions attempted to significantly update governance over the emerging area of ‘secondary research’ with data or biospecimens, i.e., research in addition to the clinical care or primary research protocol for which the biospecimen or data were originally procured.

In the wake of the immensely popular book *The Immortal Life of Henrietta Lacks,*\(^7\) and related empirical studies,\(^8\) the research community is on notice that people

---

1 Arti K. Rai, *Risk Regulation and Innovation: The Case of Rights-Encumbered Biomedical Data Silos,* 92 Notre Dame L. Rev. 1641, 1643 (2017) (‘In diagnostic testing, as in other areas of biomedicine, large data sets promote cumulative innovation.’).

2 NIH, Final NIH Policy for Data Management and Sharing (Oct. 29, 2020) https://grants.nih.gov/grants-guide/notice-files/NOT-OD-21-013.html (accessed Jan. 25, 2021) (‘Sharing scientific data accelerates biomedical research discovery, in part, by enabling validation of research results, providing accessibility to high-value datasets, and promoting data reuse for future research studies.’) (hereinafter, NIH, *Policy for data management*).

3 National Research Council, Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease (2011), https://www.nap.edu/catalog/13284/toward-precision-medicine-building-a-knowledge-network-for-biomedical-research (accessed July 17, 2020).

4 Barbara J. Evans, *Barbarians at the Gate: Consumer-Driven Health Data Commons and the Transformation of Citizen Science,* 42 Am. J. Law Med. 651, 652 (2016) (‘Data resources are a central currency of twenty-first-century science, and the question is, “Who will control them?”’).

5 45 C.F.R. § 46 (2019).

6 National Science Board, *Science & Engineering Indicators* (2018) https://www.nsf.gov/statistics/2018/nsb20181/report/sections/academic-research-and-development/expenditures-and-funding-for-academic-r-d/sources-of-support-for-academic-r-d (accessed Dec. 27, 2019).

7 Rebecca Skloot, *The Immortal Life of Henrietta Lacks* (2011).

8 Kayte Spector-Bagdady et al., *Encouraging Participation and Transparency In Biobank Research,* 37 Health Aff. (Millwood) 1313 (2018) (hereinafter Spector-Bagdady, *Encouraging Participation*).
generally want information regarding whether their specimens collected by hospitals may be ‘commercialized’ or sold to private industry—and that there are important differences in preferences by race and ethnicity.9 An updated federal regulatory requirement, consistent with this interest, stipulates that if biospecimens collected in research might later be commercialized (even if they are completely deidentified), federally-funded researchers must disclose this possibility to participants. But we also know that when people understand that their specimens may be commercialized, the majority are uncomfortable with such use.10

Also in 2019, Ascension health—which holds the medical records of patients across 20 states and the District of Columbia—shared fully identified medical records of 50 million patients with Google for research.11 Google argued that the transaction was consistent with the Business Associate Agreement requirements under the Privacy Rule of the Health Information Portability and Accountability Act (HIPAA),12 but its legality is currently under investigation by the US Department of Health & Human Service’s (DHHS) Office for Civil Rights13 and several state senators.14

Although health research has been widely recognized as a public good, an increasing amount of health data and specimens are actually under industry control, and current federal regulations have not been able to protect public accessibility. Instead, we are currently living in a myopic binary system where federally-funded researchers are highly regulated and commercial health data collectors are barely regulated at all. But Academic Medical Centers (AMCs), which receive over $12.5B in National Institutes of Health (NIH) funding per year alone,15 have more negotiation power than they are currently leveraging. Comprehensive databanks require a diversity of health-related

---

9 Reshma Jagsi et al., Perspectives of Patients with Cancer on the Ethics of Rapid-Learning Health Systems, 35 J. CLIN. ONCOL. 2315 (2017).
10 Spector-Bagdady et al., Encouraging Participation, supra note 8.
11 Ed Pilkington, Google’s secret cache of medical data includes names and full details of millions—whistleblower The GUARDIAN, Nov. 12, 2019 (‘A whistleblower who works in Project Nightingale, the secret transfer of the personal medical data of up to 50 million Americans from one of the largest health care providers in the US to Google, has expressed anger to the Guardian that patients are being kept in the dark about the massive deal.’), Tarq Shaukat, Our partnership with Ascension, https://cloud.google.com/blog/topics/inside-google-cloud/our-partnership-with-ascension (accessed Dec. 24, 2019).
12 C.F.R. § 164.514(b) (2016).
13 Rebecca Robins, HHS to probe whether Google’s ‘Project Nightingale’ followed federal privacy law STAT, Nov. 13, 2019, https://www.statnews.com/2019/11/13/hhs-probe-google-ascension-project-nightingale/ (accessed May 31, 2020) (‘A federal regulator is investigating whether the federal privacy law known as HIPAA was followed when Google collected millions of patient records through a partnership with nonprofit hospital chain Ascension . . . . “OCR would like to learn more information about this mass collection of individuals’ medical records with respect to the implications for patient privacy under HIPAA,” Roger Severino, the office’s director, said in a statement to STAT.’).
14 Heather Landi, Google defends use of patient data on Capitol Hill among scrutiny of Ascension deal, FIERCE HEALTHCARE (Mar. 4, 2020) https://www.fiercehealthcare.com/tech/senators-pressing-ascension-on-google-data-deal-as-tech-giant-defends-its-use-patient-records (accessed Feb. 13, 2021) (‘The lawmakers—presidential candidate Elizabeth Warren (D-Mass.), Richard Blumenthal (D-Conn.), and Bill Cassidy (R-La.)—sent a letter to Ascension . . . . demand[ing] more information regarding the type and amount of information the health system has provided to Google, whether the health system provided advance notice to patients about the deal and whether patients can opt-out of data sharing.’).
15 Blue Ridge Institute for Medical Research, Ranking Tables of NIH Funding to US Medical Schools in 2020 as compiled by Robert Roskoski Jr. and Tristram G. Parslow http://www.brimr.org/NIH_Awards/2020/defaul t.htm (accessed Feb. 14, 2021).
data, which is easier to collect when people have a compelling basic interest in sharing it (ie at a hospital or clinic), as well as when insurance is reimbursing to annotate health information in detail. In addition, industry needs academic collaborators to conduct research, publish articles, and treat the patients and write the prescriptions that make the private data valuable in the first place. AMCs could use this captive audience not just to negotiate for licenses for future machine-learning products (as some are currently doing) but also to protect and improve the treatment of their patients and participants.

Part I of this article will discuss the value of secondary research with data and specimens. Part II will analyze why AMCs have been put at a disadvantage as compared to industry in amassing the large datasets that enable this work. It will specifically focus on the origin of the human subjects research regulations, their scope, regulatory alternatives for industry dataset governance, and the increasing mixture of health data and specimens across entities. Part III will explore the limitations of this current governance structure, arguing that the federal regulations are not accomplishing their stated goal of informed consent, and they are also not appropriately calibrating to the risks and benefits of secondary research. In addition, the regulations as written are serving to actually encourage a private data and specimen market with associated limitations. And Part IV will propose that, moving forward, AMCs should set their own standards for commercialization of the data and specimens they generate in-house, the ability of their researchers to use industry data for their own work, and baseline informed consent standards for their own patients and participants.

The human subjects research governance structure was created to govern research with humans. It was not designed to govern research with all the stuff derived from them. The consequences of this imbalance between emerging research protocols and static regulations is the government-enabled privatization of one of our most valuable health resources. But there are steps that AMCs can and should currently take to better control the revolving door of data and specimens between AMCs and industry to protect access to the data and specimens needed for life-saving research moving forward.

II. HEALTH DATA AND BIOSPECIMENS AS A COMMON GOOD
In order to achieve promises made by the Precision Medicine Initiative and other cutting-edge health campaigns, researchers need data—regarding lifestyle choices, environment, health outcomes, and genetics—derived from biospecimens, medical records, wearable technologies, and other collection methods. Through 'big data' researchers can slowly piece together what health outcomes individuals can control, or clinicians can treat, and subsequently improve. This has led some scholars to

16 Dinerstein v. Google, LLC, No. 19 C 4311, 2020 WL 5296920, 1079 (N.D. Ill. Sept. 4, 2020) (‘Whatever a perpetual license for “Trained Models and Predictions” actually means, it appears to qualify as direct or indirect remuneration.’).

17 Anne Cambon-Thomsen, The Social and Ethical Issues of Post-Genomic Human Biobanks, 5 NAT. REV. GENET. 866, 867 (2004) (‘large population biobanks are indeed useful tools not only as a repository of genomic knowledge but also as a means of measuring non-genetic environmental factors. As such, they give epidemiologists and geneticists a new tool to explore complex gene–gene and gene–environment interactions at the population level.’).
categorize health data as critical ‘infrastructural resource’; its value derived from downstream uses rather than as an end product in and of itself.\textsuperscript{18}

Ruth Faden and colleagues have argued that medical centers should be reconceptualized as ‘learning health systems’ which would ‘have continuous access to information about as many patients as possible to be efficient, affordable, fair, and of highest quality.’\textsuperscript{19} But data and biospecimens currently used for secondary research often come from broader sources than a single medical center.\textsuperscript{20} Personal health data are collected across the Internet, apps, and other data-capture mechanisms via algorithmic systems in people’s devices, wearables, homes, work, personal lives, and leisure activities.\textsuperscript{21} This makes big data a big business.\textsuperscript{22} In 2017, the personal data market was valued at $35 billion—and it is projected to reach $103 billion by 2027.\textsuperscript{23}

Resources that are beneficial to a large part of the population, such as natural resources like fresh water or man-made resources such as data or biobanks, are often described as ‘common resources’. A unique set of theories regarding the best ways to govern these resources has direct applicability to the health data and biospecimen commons.\textsuperscript{24} An oft-invoked example of a commons is a shared pasture upon which sheep may graze. If sheep overgraze, the grass will diminish, and no sheep will be able to graze. Thus, although any individual farmer might benefit from adding to her flock and allowing an additional sheep to graze, if \textit{all} the farmers acted in such a way indefinitely, there would be no grass left for any sheep and everyone will be worse off. As Garrett Hardin\textsuperscript{25} originally argued:

\begin{quote}
Therein is the tragedy. Each man is locked into a system that compels him to increase his herd without limit – in a world that is limited. Ruin is the destination toward which all men rush, each pursuing his own best interest in a society that believes in the freedom in the commons. Freedom in a commons brings ruin to all.\textsuperscript{26}
\end{quote}

\textsuperscript{18} Brett M. Frischmann, \textit{Infrastructure: The Social Value of Shared Resources} 61 (2012) (‘[s]ocial demand for the resource is driven primarily by downstream productive activities that require the resource as an input.’); also see eg W. Nicholson Price II, \textit{Risk and Resilience in Health Data Infrastructure}, 16 COLO. TECH. L.J. 65, 77 (2017) (‘Roads are not valuable principally because you can drive on them; roads are valuable because you can use them to get places and transport goods.’).

\textsuperscript{19} Ruth R. Faden et al., \textit{An Ethics Framework for a Learning Health Care System: A Departure from Traditional Research Ethics and Clinical Ethics}, HASTINGS CENT. REP. S16–27, S23 (2013).

\textsuperscript{20} Jodyn Platt et al. \textit{Ethical, Legal, and Social Implications of Learning Health Systems}, 2 LEARN. HEALTH SYST. e10051 (2018).

\textsuperscript{21} Evans, supra note 4, at 651.

\textsuperscript{22} Elizabeth R. Pike, \textit{Defending Data: Toward Ethical Protections and Comprehensive Data Governance}, 69 EMORY L.J. 687, 691 (‘Personal data are big business.’).

\textsuperscript{23} Kari Paul, \textit{What Is Exactis—and How Could It Have Leaked The Data Of Nearly Every American?}, \textit{Market Watch}, June 29, 2018, \url{https://www.marketwatch.com/story/what-is-exactis-and-how-could-it-have-the-data-of-nearly-every-american-2018-06-28} (accessed June 29, 2018).

\textsuperscript{24} Amy L. McGuire et al., \textit{Importance of Participant-Centricity and Trust for a Sustainable Medical Information Commons}, 47 J. LAW MED. ETHICS 12, 15 (2019).

\textsuperscript{25} It is critical to note that, although Hardin’s conceptual framework of a commons founded a theory upon which many scholars have built, Hardin was a racist and eugenicist. His words should only be read and understood within that important context. Eg Craig Straub, \textit{Living in a world of limits} (interview with Garrett Hardin), 8 THE SOCIAL CONTRACT 1 (1997) (‘…I think there are other reasons for restricting immigration that are more powerful. My position is that this idea of a multiethnic society is a disaster…[it] is insanity. I think we should restrict immigration for that reason.’).

\textsuperscript{26} Garrett Hardin, \textit{The tragedy of the commons}, 162 SCIENCE. 1243, 1244 (1968).
But how to avoid such ruin? Hardin offers two solutions: privatization or government regulation (‘These, I think, are all the reasonable possibilities. They are all objectionable. But we must choose – or acquiesce in the destruction of the commons . . . .’).\(^\text{27}\) In other words, if the government is not governing a commons adequately, private interests will. Thus, under social contract theory, communities are incentivized to give up some amount of their liberty in exchange for living within a system that offers centralized protection for community well-being. But, although private industry ‘may well contribute to the common good’ they ‘are not guardians of this common good.’\(^\text{28}\) The common good is not a private matter, it is the purview of government.\(^\text{29}\) So then what should we do as a society if the government is not adequately protecting it?

### III. CURRENT GOVERNANCE OF HEALTH DATA AND BIOSPECIMENS

In order to fully discuss potential solutions to balancing the breadth and value of industry versus AMC secondary research banks, we must first understand how they are currently governed. This section will review the current human subjects research regulatory structure, discuss limitations inherent therein, explore alternative regulations potentially applicable to industry databanks, and end by arguing that—while the entire system is founded on controlling the entity that acquired the health data or specimen in the first place—data and specimens are becoming increasingly mixed across entities, thereby undermining a central premise of the entire governance system.

#### II.A. Human Subjects Research Regulations

The current US human subjects research regulations were developed in the 1970s and ‘80s in the wake of research catastrophes, such as the infamous syphilis experiments in Tuskegee, Alabama conducted by the US Public Health Service.\(^\text{30}\) These regulations are steadfastly founded in the need for the informed consent of the individual,\(^\text{31}\) as a demonstration of her autonomy interests, before she may be enrolled in research as a participant. Case law focused on secondary research use of data and specimens, however, narrowly assesses them on their value to research—as opposed to the individual from whom they were derived. However, new cases regarding modern data-sharing relationships between AMCs and industry may begin to elucidate how the law will attempt to control these kinds of relationships moving forward.

#### III.A.1. The Need to Regulate Human Subjects Research

Starting with the Hippocratic Oath, the original foundation of medical ethics as a field rested in the concept of beneficence: the idea that clinicians ought not to inflict harm upon patients and should instead promote good.\(^\text{32}\) Even after the atrocities committed during the Holocaust by Nazi doctors and ‘researchers’, as well as reckonings by Henry Beecher and other critics of the American research enterprise closer to

\(^{27}\) Id.

\(^{28}\) Jonathan H. Marks, The Perils of Partnership, at 34 (2019) (emphasis added).

\(^{29}\) Id.

\(^{30}\) James H Jones, Bad Blood: The Tuskegee Syphilis Experiments (1993).

\(^{31}\) Or surrogate, if the individual lacks capacity.

\(^{32}\) Hippocrates, The Oath (Francis Adams trans. 400 B.C.), http://classics.mit.edu/Hippocrates/hippooath.html [http://perma.cc/B2JH-86RE] (accessed July 17, 2020).
home, there was still an overarching sentiment by the federal government that US clinical researchers were upstanding citizens, just like their clinical counterparts, and their motivations should generally not be questioned.

This assumption was challenged in the 1970s with the public revelation of the now-infamous syphilis experiment in Tuskegee, Alabama. During that experiment, researchers from the US Public Health Service lied to impoverished African American men with syphilis, whom they went on to study without notice or consent. In 1932, when the syphilis experiments in Tuskegee began, the only known treatments for syphilis (eg arsenic or mercury) required a long course, were largely ineffective, and had many adverse effects. However, in 1943, the same team of Public Health Service researchers discovered that penicillin could effectively cure syphilis.

The syphilis experiment in Tuskegee then took an even darker turn when researchers began attempting to prevent subjects from accessing penicillin that might treat their syphilis . . . for the next 30 years. It was not until an internal US Centers for Disease Control and Prevention whistle-blower and a journalist brought the syphilis experiment in Tuskegee to the attention of the US media in 1972 that the study was finally stopped. In addition, once discovering a cure for syphilis, many of the same researchers turned to improving prophylactic measures to prevent it in the first place, and conducted experiments between 1946–48 in Guatemala for which they purposefully infected vulnerable subjects with STDs.

---

33 Henry K. Beecher, *Ethics and Clinical Research*, 274 N. ENGL J. MED. 1354 (1966) (‘Human experimentation since World War II has created some difficult problems with the increasing employment of patients as experimental subjects when it must be apparent that they would not have been available if they had been truly aware of the uses that would be made of them. Evidence is at hand that many of the patients in the examples to follow never had the risk satisfactorily explained to them, and it seems obvious that further hundreds have not known that they were the subjects of an experiment although grave consequences have been suffered as a direct result of experiments described here.’).

34 Kayte Spector-Bagdady & Paul A. Lombardo, ‘Something of an adventure’: postwar NIH Research Ethos and the Guatemala STD Experiments 41 J. LAW MED. ETHICS 697 (2013).

35 RUTH R. FADEN & TOM L. BEAUCHAMP, A HISTORY AND THEORY OF INFORMED CONSENT, at 23 (1986); Tom L. Beauchamp, *Informed Consent: Its History, Meaning, And Present Challenges*, 20 CAMBRIDGE Q. HEALTHCARE ETHICS 515, 518 (2011).

36 H. A. Callis, *Comparative Therapy in Syphilis*, 21 J. NATL. MED. ASSOC. 61 (1929).

37 John F. Mahoney et al., *Penicillin Treatment of Early Syphilis: A Preliminary Report*, 33 AM. J. PUBLIC HEALTH & NATION’S HEALTH 1390 (1943).

38 In many cases this was not actually successful, making the Tuskegee experiments more of a study of under-rather than untreated syphilis. Susan M. Reverby, *Compensation and Reparations for Victims and Bystanders of the U.S. Public Health Service Research Studies in Tuskegee and Guatemala: Who Do We Owe What?* BIOETHICS DOI: 10.1111/bioe.12784 (2020) (‘. . . some of the men still alive in the post World War II antibiotic era were able to get to treatment, sometimes because they had moved outside of the area, or because their doctors did not know they were in the study.’).

39 Jones, supra note 30.

40 Presidential Commission for the Study of Bioethical Issues, Ethically Impossible: STD Research in Guatemala from 1946 to 1948 (2011) https://bioethicsarchive.georgetown.edu/pcsbi/sites/default/files/Ethically%20Impossible%20(with%20linked%20historical%20documents)%202.7.13.pdf (accessed July 18, 2020).
were kept secret until rediscovered in 2010. The US clinical research system clearly suffered from critical flaws in oversight with terrible consequences.

III.A.2. The Common Rule

In the face of such interventional research failures—with risks and burdens to subjects that were individual, physical, and profound—US bioethicists and regulators reconsidered treating clinical researchers with the same deference given to clinicians. Instead of presuming that the motivation of a researcher was beneficence, regulators instead began to emphasize the potential conflict of interest a researcher has with promoting her own work. Under this view, proposed research protocols had to be assessed by neutral third parties in order to protect the participant from the conflicted researcher. The role of the third party would be to ensure that any potential risks of research were counterbalanced by potential benefits (either to the individual or society) and that individual subjects provided fully informed consent before enrollment. The Belmont Report, authored by the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research in 1979, thus cemented autonomy’s preeminence as the cornerstone of clinical research ethics.

Ironically, although the syphilis experiment in Tuskegee was the inciting incident for the Belmont Report, one of the framework’s main takeaways—that participants should give informed consent to participation in research—would not actually have resolved the profound problems with that work. Although researcher deceit in Tuskegee was surely a great moral defect, even if the participants had consented the experiment should never have taken place. There was also irreconcilable beneficence and justice deficits (the other two Belmont principles) as the potential benefits of the study were not proportional to the risks undertaken and the study targeted impoverished and uneducated Black participants for research that was not intended to specifically help them or their community.

But there is a policy advantage in the regulations’ focus on informed consent. Generating a list of consent form disclosure requirements to be presented to prospective enrollees is an easier goal to achieve than writing the rules that somehow assure justice and fairness. The government thus codified the predominance of disclosure, framed as

41 Susan M. Reverby, ‘Normal Exposure’ and Inoculation Syphilis: A PHS ‘Tuskegee’ Doctor in Guatemala, 1946–1948, 23 J. Policy Hist. 6 (2011).
42 Kayte Spector-Bagdady & Paul A. Lombardo, U.S. Public Health Service STD Experiments in Guatemala (1946–1948) and Their Aftermath, 41 Ethics Hum. Res. 29 (2019).
43 NATIONAL BIOETHICS ADVISORY COMMISSION, THE BELMONT REPORT (1979) https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/read-the-belmont-report/index.html#xrespect (accessed July 18, 2020) (‘In most cases of research involving human subjects, respect for persons demands that subjects enter into the research voluntarily and with adequate information.’); Jonathan Beever & Nicolae Morar, The Porosity of Autonomy: Social and Biological Constitution of the Patient in Biomedicine, 16 Am. J. Bioeth. 34 (2016) (‘Respect for the individual holds the place of utmost prominence among the principles of contemporary bioethics.’).
44 Ruqaiijah Yearby, Exploitation in Medical Research: The Enduring Legacy of the Tuskegee Syphilis Study, 67 Case W. Res. 1171, 1175 (2017).
45 For an argument regarding why principlism is oversimplified altogether, see John H. Evans, A Sociological Account of the Growth of Principlism, 30 Hastings Cent. Rep. S, 31–38 (2000) (‘Principlism is . . . a method that takes the complexity of actually lived moral life and translates this information into four scales by discarding information that resists translation.’).
supporting the informed consent necessary to enable autonomy, into the foundation of the human subjects research governance structure. The first subpart of these regulations were later adopted across the majority of the US federal departments and agencies and thus named the ‘Common Rule’.  

The Common Rule generally requires Institutional Review Board (IRB) review of proposed protocols to ensure there is an appropriate balance between risks and burdens to participants and potential benefits to the participant herself, or to knowledge generally. There are additional protections for vulnerable populations, such as children, in subsequent subparts. If an IRB has authorized recruitment of participants into a study, individual consent must be given by the individual participant or her representative. And detailed consent form disclosures—such as information about risk, benefits, and alternatives—now act as the gatekeepers through which all generalizable knowledge must pass. Considerations of justice and beneficence are left to the mercies of individual IRBs to assess on a case by case basis. The Common Rule had remained substantially similar to its first iteration until 2019 when it was revised after a 6-year regulatory notice and comment process.

III.A.3. Case Law

Although regulatory protections for human subjects research has largely focused on the scope of which participants should be protected, much of the famous human subjects research case law has focused on use and access to biospecimens via the classic trio of Moore v. Regents, Greenberg v. Miami, and Washington University v. Catalona. Together, these cases are less deferential to the autonomy interests of the people from whom the specimens are derived, and are instead founded in supporting researcher access to valuable biobanks. The new dismissal in Dinerstein v. Google demonstrates how some of these issues surrounding data sharing specifically might be litigated henceforth.

In Moore v. Regents (1990), patient John Moore went to doctors at the University of California, Los Angeles for the treatment of his hairy-cell leukemia. His clinicians collected many types of biospecimens from him, including samples taken after such collection was no longer necessary for his clinical care, without disclosing that the purpose of the ongoing collection was research. The investigators were ultimately able...
to develop a lucrative cell line from his specimens. Once Moore realized what his doctors were doing, he sued for conversion (among other things).56

The California Supreme Court found that a claim of conversion could not stand because Moore did not retain an ownership interest in his cells once they left his body. It did find that the doctors should have disclosed their research interest to Moore as a potential conflict of interest during his clinical informed consent process.57 The court ultimately held that ‘the theory of liability that Moore urges us to endorse threatens to destroy the economic incentive to conduct important medical research . . . . If the use of cells in research is a conversion, then with every cell sample a researcher purchases a ticket in a litigation lottery.’58

In Greenberg v. Miami (2003),59 parents of children affected by Canavan disease, a devastating neurodegenerative disorder, donated money, medical information, and biospecimens to researchers at Miami Children’s hospital. The families stated that their goal was to support the research necessary to isolate the variants associated with (and thereby develop a carrier test for) the disease, as well as work toward a cure. The Miami researchers did in fact identify the variant associated with Canavan, but patented the discovery.60 The patent restricted others’ ability to offer Canavan carrier testing or do their own therapeutic research.

The families sued for claims including a lack of informed consent and conversion based on their understanding that their contributions were in exchange for ‘affordable and accessible’ testing for other families.61 The district court in this case declined to recognize a failure in the duty of informed consent between the families and the researchers. The Greenberg court distinguished Moore due to the ‘therapeutic relationship’ between Moore and his physician researchers; in Greenberg the court described defendants as ‘solely medical researchers.’62 The court bemoaned that finding otherwise would ‘have pernicious effects over medical research . . . .’63 The conversion claim was also dismissed since ‘the property right in blood and tissue samples also evaporates once the sample is voluntarily given to a third party.’64

In the last famous biospecimen case, Washington University v. Catalona (2006),65 Dr William Catalona planned to move his faculty practice and research on prostate cancer from Washington University to Northwestern. He had personally recruited ∼3000 of his patients to contribute samples to the WashU biobank.66 In anticipation of his move,

56 Jessica L. Roberts, Genetic Conversion, Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3357566 (accessed July 16, 2020).
57 Moore, S1 3d at 131 (‘Accordingly, we hold that a physician who is seeking a patient’s consent for a medical procedure must, in order to satisfy his fiduciary duty and to obtain the patient’s informed consent, disclose personal interests unrelated to the patient’s health, whether research or economic, that may affect his medical judgment.’).
58 Moore, S1 3d at 146.
59 Greenberg, 264 F. Supp. 2d.
60 The Supreme Court subsequently found that a patent on a genetic variant is invalid “merely because it has been isolated”, see Association for Molecular Pathology v. Myriad Genetics, Inc., 569 U.S. 576 (2013).
61 Greenberg, 264 F. Supp. 2d at 1067.
62 Id., at 1070.
63 Id.
64 Id., at 1075.
65 Wash. Univ., 490 F.3d.
66 Id. at 988–89.
Dr Catalona sent a letter to the much broader cohort of all (∼30,000) contributors whose specimens had been used in his research asking them to donate and release their samples to him at Northwestern for use ‘only at his discretion and with his express consent . . . .’. Six thousand agreed.\textsuperscript{67} WashU sued.

The Missouri district court found for WashU in part with reference to Moore and Greenberg,\textsuperscript{68} and in so doing made an impassioned policy argument for the importance of enabling accessible biobanks:

Medical research can only advance if access to these materials to the scientific community is not thwarted by private agendas. If left unregulated and to the whims of a [research participant], these highly prized biological materials would become nothing more than chattel going to the highest bidder. It would no longer be a question of the importance of the research protocol to public health, but rather who can pay the most. [ . . . ] The integrity and utility of all biorepositories would be seriously threatened if [research participants] could move their samples from institution to institution any time they wanted. No longer could research protocols rely on aggregate collections since individual samples would come and go.\textsuperscript{69}

A more recent example related to data-sharing agreements between an AMC and industry comes in the form of \textit{Dinerstein v. Google}\textsuperscript{70} from the Northern District of Illinois. At issue in \textit{Dinerstein} is a 2017 data use agreement between the University of Chicago and Google. The goal of this agreement was to generate ‘machine-learning techniques to create predictive health models aimed at reducing hospital readmissions and anticipating future medical events’.\textsuperscript{71} To do so, UChicago shared the ‘de-identified’ electronic medical records (EMR) of all adult patients over a 5-year period (the data, in fact, included dates of service).\textsuperscript{72}

When Mr Dinerstein, a patient, found out his health data had been shared with Google, he sued for breach of contract (among other things)\textsuperscript{73} for alleged violations including HIPAA compliance (as the court allowed a private right of action for HIPAA under a state tort law claim).\textsuperscript{74,75} Under HIPAA, covered entities may not generally sell protected health information (PHI) without written permission. But this prohibition does not include sharing PHI for research purposes in exchange for a ‘reasonable cost based fee to cover the cost to prepare and transmit’ said PHI.\textsuperscript{76} In \textit{Dinerstein}, the data use agreement between Google and UChicago granted the University, for internal ‘non-commercial research’ purposes, ‘a nonexclusive, perpetual license to use the [ ] Trained

\textsuperscript{67} Id. at 993.
\textsuperscript{68} Id. at 995–97 (‘The two cases which provide the most guidance conclude that research participants retain no ownership of biological materials they contribute for medical research.’).
\textsuperscript{69} Id. at 1000.
\textsuperscript{70} Dinerstein v. Google, LLC, No. 19 C 4311, 2020 WL 5296920, 1079–1124 (N.D. Ill. Sept. 4, 2020); see also I. Glenn Cohen & Michelle M. Mello, \textit{Big Data, Big Tech, and Protecting Patient Privacy}, 322 JAMA 1141 (2019) (hereinafter Cohen & Mello, \textit{Big Data}).
\textsuperscript{71} Dinerstein, 2020 WL 5296920 at 1079.
\textsuperscript{72} Id. at 1079.
\textsuperscript{73} Dinerstein, 2020 WL 5296920 at 1099.
\textsuperscript{74} Id. at 1104.
\textsuperscript{75} Id. at 1112.
\textsuperscript{76} 45 C.F.R. §§ 164.502(a)(5)(ii)(B)(2)(ii).
Models and Predictions’ created by Google.\textsuperscript{77} Defendants argued that the perpetual license was, in fact, only a reasonable cost-based fee. However the Court found that: ‘Whatever a perpetual license for “Trained Models and Predictions” actually means, it appears to qualify as direct or indirect remuneration.’\textsuperscript{78} But ultimately the Court ruled that none of Mr Dinerstein’s arguments could support a claim for relief\textsuperscript{79} because Illinois neither recognizes noneconomic breach of contract damages, eg ‘anxiety and distress’ (but for under exceptional circumstances),\textsuperscript{80} nor could the plaintiff establish economic damages to his EMR data (which were not even recognized as his property to begin with).\textsuperscript{81} In summary, it will be exceptionally hard for plaintiffs to move past dismissal in the future unless they can establish that the invasion of their privacy resulted in specific financial injury.

III.B. Limitations in the Scope of the Human Subjects Research Regulations
Despite the rich regulatory and case law background of the governance of human subjects research, a major limitation is that this governance structure does not actually cover all human subjects research. Even where the regulations do apply, they only govern research that is interventional or involves identified biospecimens or data—which excludes a large number of secondary protocols. And, even when the regulations apply, and even if the biospecimens or data are identifiable, there are still options for conducting the research with a waiver of informed consent. As Neil Richards and Woodrow Hartzog have argued: consent ‘transforms the moral landscape between people and makes the otherwise impossible possible.’\textsuperscript{82} But the human subjects research regulations allow an exceptional number of exceptions to actually procuring such consent.

III.B.1. Not All Human Subjects Research
First, the human subjects research regulations do not actually apply to all human subjects research. There are, in fact, only four circumstances in which they do apply: (i) If the investigator is conducting research using federal funding from a US Department or Agency that has adopted the Common Rule.\textsuperscript{83} This is understood as a reasonable derivation of Congressional spending power, ie if you are going to take the government’s money to do research, it has the right to put limits on usage.\textsuperscript{84} When the Common Rule was revised in 2018, regulators considered extending its reach to include all clinical

\textsuperscript{77} Dinerstein, 2020 WL 5296920 at 1082.
\textsuperscript{78} Id. at 1109.
\textsuperscript{79} Id. at 1119.
\textsuperscript{80} Parks v. Wells Fargo Home Mortg., Inc., 398 F.3d 937, 940–41 (7th Cir. 2005) (quoting Maere v. Churchill, 116 Ill. App. 3d 939, 944, 452 N.E.2d 694, 697 (3d Dist. 1983)).
\textsuperscript{81} Dinerstein, 2020 WL 5296920 at 1118.
\textsuperscript{82} Neil Richards & Woodrow Hartzog, The Pathologies of Digital Consent, 96 Wash. U. L. Rev. 1461, 1462 (2019).
\textsuperscript{83} 45 C.F.R. § 46.101 (2019) (‘…this policy applies to all research involving human subjects conducted, supported, or otherwise subject to regulation by any Federal department or agency that takes appropriate administrative action to make the policy applicable to such research.’).
\textsuperscript{84} Helvering v. Davis, 301 U.S. 619, 645 (1937) (‘…when money is spent to promote the general welfare, the concept of welfare or the opposite is shaped by Congress …’); Institute of Medicine, Committee on Ethical Considerations for Revisions to DHHS Regulations for Protection of Prisoners Involved in Research (Gostin LO, Vanchieri C, Pope A, eds.) (2007).
trials at US institutions that receive some federal support for human subjects research (regardless of the funding of the specific study). But the final iteration of the updated rule did not extend its scope. (ii) If an institution voluntarily decides to extend the regulatory requirements to all of its employees conducting human subjects research (as many do). (iii) If researchers are using an investigational-only product which requires US Food and Drug Administration (FDA) authorization to ship and use in interstate commerce. Or (iv), if investigators wish to submit data derived from their research to FDA in support of an application for research or marketing, they have to follow substantially similar FDA regulations regarding the protection of human subjects.

These limitations mean that privately funded research is outside the scope of federal regulations if it does not involve product requiring FDA authorization to distribute or if researchers do not ultimately submit their data in support of an FDA application. Some private entities may choose to follow components of the research regulatory structure due to other market motivations, discussed further below, but many do not. Recent more comprehensive data privacy legislation has been adopted in Europe and California, although both still only regulate ‘identified’ data, have ambiguously broad exceptions for research in public health, and do not cover health-related data sharing in the USA. For example under the GDPR, although participants in clinical trials must give full informed consent for future data sharing (‘organization employed clear, intelligible, and easily accessible language, with the purpose for data processing

85 Department of Homeland Security et al., Notice of Proposed Rulemaking: Federal Policy for the Protection of Human Subject. 80 Fed. Reg. 173, 53933–54061, 53989 (Sept. 8, 2015).
86 This is due to the fact that regulators agreed with the ‘slim majority’ of commenters opposing the change and ultimately agreed that such an extension would ‘benefit from further deliberation’. Department of Homeland Security et al., Federal Policy for the Protection of Human Subjects 82 Fed. Reg. 12, 7149–7274, 7155–56 (Jan. 19, 2017).
87 Id. at 7156 (‘We recognize that institutions may choose to establish an institutional policy that would require IRB review of research that is not funded by a Common Rule department or agency (and indeed, as commenters noted, almost all institutions already do this), and nothing in this final rule precludes institutions from providing protections to human subjects in this way.’), although the revisions also did away with the previous ‘Federal Wide Assurance’ mechanism under which institutions could contractually commit to the government that they would do so (‘We therefore plan to implement the proposed nonregulatory change to the assurance mechanism to eliminate the voluntary extension of the FWA to nonfederally funded research.’).
88 Eg 21 C.F.R. §312 (1987).
89 21 CFR § 50.1 (2019).
90 Presidential Commission for the Study of Bioethical Issues, Moral Science: Protecting Participants in Human Subjects Research (2011) https://bioethicsarchive.georgetown.edu/pcsbi/node/$58.html (accessed Dec. 29, 2019).
91 Kayte Spector-Bagdady et al., Nonregulated Interventions, Clinical Trial Registration, and Editorial Responsibility, 12 Circ. Cardiovasc. Qual. Outcomes E005721 (2019) (‘Because of inherent limitations of the scope of enforcement by FDA, funders, and research institutions, a critical fourth wall of protection can also be journal standards. Expectations set by journal editors can influence major components of the international research industry.’).
92 General Data Protection Regulation 2016/679 (2018).
93 California Consumer Privacy Act, AB No. 375 (2018).
94 W. Nicholson Price II et al., Shadow Health Records Meet New Data Privacy Laws, 363 Science 448, 450 (2019) (‘the GDPR refers to exceptions for “scientific research”, the “public interest”, and “public health” without clearly defining these overlapping terms or addressing dual-use endeavors.’) (hereinafter Price et al., Shadow health records).
attached to that consent\textsuperscript{95}, and they must be able to withdraw from future trials, they cannot retrospectively erase their clinical data without an audit trail.\textsuperscript{96}

In addition, and somewhat counter-intuitively, research with health data and specimens collected via clinical care from patients is also not governed by the human subjects research regulations. This work instead falls under the HIPAA research rules. But HIPAA only covers individuals’ PHI collected by ‘covered entities’ (ie health care providers, health plans, or healthcare clearing houses). Covered entities are allowed to share identified data with ‘business associates’ with whom they have a contract to perform ‘functions or activities on behalf of’ or provide services.\textsuperscript{97} HIPAA allows entities to transfer data for research as long as they are deidentified,\textsuperscript{98} but also allows for the research sharing of identified data and specimens with an IRB waiver—a much more efficient process than acquiring consent in the first place.\textsuperscript{99} And the mandatory disclosures regarding default consent to data and specimens collection for research is generally written right into an institution’s standard clinical consent form.\textsuperscript{100} As W. Nicholson Price II has pointed out:

‘. . . if health privacy is worth defending, then why limit those defenses to the narrow set of actors and data covered by HIPAA, as the United States largely does? HIPAA’s outdated focus on covered entities and its safe harbor for “deidentified” data leave too much for manipulation, if health privacy protection is the goal.’\textsuperscript{101}

But, as tested in \textit{Dinerstein v. Google},\textsuperscript{102} built into HIPAA is the assumption that private health information is only collected by covered entities, or, that once collected, those data stay put.\textsuperscript{103,104} In addition, HIPAA neither envisioned nor protects the huge swath of health-proxy data from which entities can derive health-related information

\textsuperscript{95} Mabel Crescioni & Tara Sklar, \textit{The Research Exemption Carve Out: Understanding Research Participants Rights Under GDPR and U.S. Data Privacy Laws}, 60 JURIMETRICS 125, 128 (2020).
\textsuperscript{96} Id. at 128 (‘Exemptions included in the law allow sponsors to refuse the request for the data to be removed, but this exemption has yet to be interpreted or applied by the courts.’).
\textsuperscript{97} Dept. Health & Human Services, \textit{Business Associate Contracts}, Jan. 25, 2013, \url{https://www.hhs.gov/hipaa/for-professionals/covered-entities/sample-business-associate-agreement-provisions/index.html} (accessed Dec. 25, 2019).
\textsuperscript{98} 45 C.F.R. § 164.514 (b) (2016).
\textsuperscript{99} W. Nicholson Price II, \textit{Problematic Interactions between AI and Health Privacy}, \textit{UTAH L. REV.} (forthcoming) (hereinafter Price, \textit{Problematic interactions}).
\textsuperscript{100} Kayte Spector-Bagdady, \textit{Hospitals Should Act Now to Notify Patients About Research Use of Their Data and Biospecimens} 26 NAT. MED. 306 (2020).
\textsuperscript{101} W. Nicholson Price II, \textit{Medical AI and Contextual Bias}, 33 Harv. J.L. & Tech. 65, *** (2020) (hereinafter Price, \textit{Contextual bias}).
\textsuperscript{102} \textit{Dinerstein}, 2020 WL 5296920 at 1079.
\textsuperscript{103} W. Nicholson Price II & I. Glenn Cohen, \textit{Privacy in the Age of Medical Big Data}, 25 NAT. MED. 37, 39 (2019) (‘When Congress enacted HIPAA in 1996, it envisioned a regime in which most health data would be held in health records, and so it accordingly focused on health care providers and other covered entities. In the big-data world, the type of data sources covered by HIPAA are but a small part of a larger health data ecosystem.’).
\textsuperscript{104} Stacey A. Tovino, \textit{Assumed Compliance}, 72 ALA. L. REV. 279, 282 (2020) (‘The HIPAA Rules do not protect the privacy and security of health data collected, used, disclosed, or sold by many technology companies, online service providers, mobile health applications, and other entities and technologies that do not meet the definition of a covered entity or business associate.’).
(eg Google’s non-UChicago derived health information).\textsuperscript{105} This might include, for example, geolocation data about visits to a psychiatrist, or demographic information often correlated with health outcomes such as maternal mortality.\textsuperscript{106}

### III.B.2. Readily Identifiable Health Data and Biospecimens Only

A second important limitation of the research regulations is that—even if research is federally funded, at an institution that requires all researchers to follow them, or generated with or as part of an FDA application—the rules only apply to research involving ‘human subjects’.\textsuperscript{107} A ‘human subject’, in turn, is defined as a living person with whom the investigator ‘obtains information through intervention or interaction’ or which involves ‘identifiable private information or identifiable biospecimens’.\textsuperscript{108} Thus, whether a biospecimen or data are considered ‘identified’ remains a critical threshold requirement for protections.

Past concerns regarding individual identifiability of data have focused on large-scale or germline genomic data. As such genetic sequences are unique to only one person, they were assumed to be more easily reidentified than other types of deidentified health data.\textsuperscript{109} However, recent research has demonstrated that now almost all data from Americans can actually be reidentified by matching it with as few as 15 demographic attributes that are easily discoverable from the information we enter into our phones, computers, and wearables every day\textsuperscript{110} (unless you are Latanya Sweeney, in which case you only need three\textsuperscript{111}). The time when genes represented a singularly unique identifier worthy of potential additional protection has already come and gone. Thus additional laws protecting people from untoward uses of those data, ie the Genetic Information Nondiscrimination Act of 2008,\textsuperscript{112} are now far too limited in scope to

\begin{itemize}
\item \textsuperscript{105} Price, \textit{Problematic interactions}, supra note 99.
\item \textsuperscript{106} I. Glenn Cohen & Michelle M. Mello, \textit{HIPAA and Protecting Health Information in the 21st Century}, 320 JAMA 231, 232 (2018) (‘HIPAA does not cover health or health care data generated by noncovered entities or patient-generated information about health (eg social-media posts). It does not touch the huge volume of data that is not directly about health but permits inferences about health . . . . The amount of such data collected and trade online is increasing exponentially and eventually may support more accurate predictions about health than a person’s medical records.’).
\item \textsuperscript{107} 45 CFR § 46.101(a) (2019).
\item \textsuperscript{108} 45 CFR § 46.102(e) (2019).
\item \textsuperscript{109} \textit{Presidential Commission for the Study of Bioethical Issues, Privacy and Progress in Whole Genome Sequencing}, at 83 (Oct. 2012) \url{https://bioethicsarchive.georgetown.edu/pcsbi/sites/default/files/PrivacyProgress508_1.pdf} (accessed Aug. 2, 2019) (‘Obtaining a whole genome sequence data file by itself yields information about, but does not definitively identify, a specific individual. The individual still has “practical obscurity”, as his or her identity is not readily ascertainable from the data. Practical obscurity means that simply because information is accessible, does not mean it is easily available or interpretable, and that those who want to find specific information must expend a lot of effort to do so . . . . In addition, even if we know that a whole genome sequence is from one individual, we cannot know which of the over 7 billion people on Earth that person is without a key linking the whole genome sequence information with a single person or their close relative. Therefore, while whole genome sequence data are uniquely identifiable, they are not currently readily identifiable.’) (hereinafter, PCSBI, \textit{Privacy & Progress}).
\item \textsuperscript{110} Luc Rocher et al., \textit{Estimating the Success of Re-Identifications in Incomplete Datasets Using Generative Models}, 10 Nat. Commun. 3069 (2019).
\item \textsuperscript{111} Latanya Sweeney, \textit{Research accomplishments}, \url{http://latanyasweeney.org/work/identifiability.html} (accessed Nov. 13, 2021).
\item \textsuperscript{112} Public Law 110–233; Genetic Information Nondiscrimination Act (2008).
\end{itemize}
protect people from the myriad types of health data which may be linked back to them in discriminatory ways.\textsuperscript{113}

\textbf{III.B.3. Waiver of Informed Consent}

Last, it is worth noting that even if research falls within the scope of federal regulation and even if participants fall within the definition of ‘human subjects’, researchers may still apply for a waiver of informed consent. Waiver can be granted by an IRB if the research involves no more than minimal risk, could not otherwise be practically carried out, would not adversely affect the rights or welfare of participants, and if additional relevant information will be provided to subjects after their ‘participation’.\textsuperscript{114}

Thus, despite the enormous research governance system set up by the federal government; parallel research protection programs at individual research institutions; and complex review, waiver, and exception policies with which researchers must grapple; many instances remain where informed consent of individual participants is not legally required to begin with. According to Price, ‘… privacy hurdles are just that—hurdles, not walls. They can be surmounted.’\textsuperscript{115} The world of ‘aconsensual’\textsuperscript{116} secondary research is still vast.

\textbf{III.C. FDA and the Collection of Industry Health Data}

Despite not being generally covered under the human subjects research regime, the private human health data market has been influenced, however circuitously, by FDA via its involvement in the direct-to-consumer personal genetic testing (DTC-PGT) industry. Although FDA has authority over a broad swath of drug, device, and biologic products including genetic tests, it exercised ‘regulatory discretion’ over DTC-PGT industry from its inception in 2007 until 2010.\textsuperscript{117} But, in 2010, Pathway Genomics, the largest DTC-PGT player at the time, announced a partnership with Walgreens under which it would sell its product in stores across the USA (in contrast to its earlier more discrete online sales model).\textsuperscript{118} FDA assessed that this increase in access to potential consumers increased the absolute risk of the product,\textsuperscript{119} presumably because more individuals were likely to buy them and therefore more individuals would suffer related burdens. FDA therefore sent ‘Untitled Letters’ warning of potential violations of the Food, Drug and Cosmetic Act to 23 DTC-PGT companies.\textsuperscript{120}

\begin{itemize}
\item \textsuperscript{113} Ellen Wright Clayton et al., \textit{The Law of Genetic Privacy: Applications, Implications, and Limitations}, 6 J. L. BIOSCI. 11 (2019) (‘The Privacy Rule was never intended to be a comprehensive health privacy regulation, but it has assumed such a role by default because of Congress’s failure to enact more sweeping and rigorous health and genetic privacy laws and regulations.’).
\item \textsuperscript{114} 45 CFR § 46.116 (2019).
\item \textsuperscript{115} Price, \textit{Problematic interactions}, supra note 99.
\item \textsuperscript{116} Holly Fernandez Lynch, et al., \textit{Implementing Regulatory Broad Consent Under the Revised Common Rule: Clarifying Key Points and the Need for Evidence}, 47 J. LAW MED. ETHICS. 213 (2019).
\item \textsuperscript{117} Kayte Spector-Bagdady & Elizabeth Pike, \textit{Consuming Genomics: Regulating Direct-to-Consumer Genetic and Genomic Information}, 92 Neb. L. REV. 677, 698 (2014).
\item \textsuperscript{118} Andrew Pollack, \textit{Walgreens Delays Selling Personal Genetic Test Kit}, N.Y. TIMES, May 12, 2010, at B5.
\item \textsuperscript{119} Spector-Bagdady & Pike, supra note 117, at 705 (‘An FDA representative stated that the DTC distribution of genetic tests can increase the risk of a device because “a patient may make a decision that adversely affects [his or her] health, such as stopping or changing the dose of a medication or continuing an unhealthy lifestyle, without the intervention of a learned intermediary.”’)
\item \textsuperscript{120} Id. at 705–10.
\end{itemize}
Untitled Letters did not require the companies to actually discontinue their product, most companies did. 23andMe, which had deep pockets due to its heavy investment from Google, 121 was the only DTC-PGT company left standing to begin regulatory authorization.122

But 23andMe, somewhat disingenuously, continued to market its product . . . while seeking FDA authorization to market its product. By 2013, FDA sent 23andMe a ‘Warning Letter’ requiring the company to discontinue its health-related testing sales. FDA justified its response due to concern about the ‘public health consequences of inaccurate results from the device.’123 23andMe complied, withdrew its health-related testing from the market, and began to offer components again only after they became FDA authorized.

Although genetic data are but one area of health-related data making up large databanks, it is an important one. The FDA/23andMe saga is an example of how the government can effectively regulate products within its purview. However, FDA’s reach over DTC-PGT data was by proxy. What FDA actually has authority over is the genetic testing device and informational results returned to consumers, not the data or research itself. Regulating the secondary use of 23andMe’s over 12-million person genetic and phenotypic database,124 or Ancestry.com’s 18 million one,125 remains outside both the purview of FDA and DHHS.

III.D. Data and Specimens between AMCs and Industry are Becoming Increasingly Mixed

Despite the fact that federal governance structure over health data and biospecimens focuses on controlling the actions of the entity that collects the data or specimens in the first place, data and specimens are increasingly becoming mixed across clinical, research, and private entities after collection—without additional informed consent. Specifically, the sharing of data and biospecimens between academia and industry has been growing at a rapid pace.

In addition to the deal already discussed in the introduction between Ascension and Google, where identified data are flowing from a covered entity to a private entity (otherwise in the business of collecting widespread data about people) under a HIPAA business associate agreement; and the one at issue in Dinerstein v. Google, where EMR data were shared from UChicago to Google in exchange for licensing rights of future algorithmic decision-making tools;126 Google also has major deals with Mayo Clinic,127 the

---

121 Lisa Baertlein, Google-backed 23andMe Offers $999 DNA Test, USA TODAY, Nov. 20, 2007, http://usatoday30.usatoday.com/tech/webguide/internetlife/2007-11-20-23andme-launch_N.htm (accessed July 18, 2020).

122 23andMeBlog, 23andMe Takes First Step Toward FDA Clearance, July 30, 2012, http://blog.23andme.com/news/23andme-takes-first-step-toward-fda-clearance/ (accessed July 18, 2020).

123 Spector-Bagdady & Pike, supra note 117, at 705–10.

124 23andMe, About Us https://mediacenter.23andme.com/company/about-us/ (accessed July 17, 2020).

125 Erin Brodwin & Katie Palmer, 5 burning questions on the business of big genetics based on 23andMe’s filing to go public, STAT+ (Feb. 5, 2021) https://www.statnews.com/2021/02/05/23andme-public-profit-genetics-data/ (accessed Feb. 13, 2021).

126 Dinerstein, 2020 WL 5296920 at 1079.

127 Casey Ross, Google, Mayo Clinic strike sweeping partnership on patient data, STAT NEWS, Sept. 12, 2019 https://www.statnews.com/2019/09/10/google-mayo-clinic-partnership-patient-data/ (accessed July 18, 2020) (‘Mayo Clinic, one of medicine’s most prestigious brands, announced Tuesday that it has
University of California, San Francisco (UCSF), and Stanford Medicine. Note that all of these deals actually involve patient data (i.e., not research participant data). Therefore, they are not primarily analyzed under the human subjects research regulations, but rather the more flexible HIPAA research regime discussed above.

But, even if medical record information that is shared is fully deidentified, and therefore does not require a HIPAA business associate agreement (which would also include other prescriptions on use), with the amount of personal data Google has already collected regarding the average person, it would be fairly straightforward to reidentify someone’s data and link their identity back to their medical information. Mr. Dinerstein attempted to use that argument in his case against Google (i.e., that there was a higher than average risk of reidentification of his EMR data by Google because of ‘the information Google already possessed about individuals through the other services it provides’), but the Northern District of Illinois Court agreed with Google that the theoretical ability to reidentify the data, in the absence of evidence that it had actually done so, was not enough to demonstrate a violation. Of note, the Google/UCSF agreement does specifically allow Google to connect the otherwise deidentified health data with other ‘data and materials it obtains elsewhere.’ Thus, ironically, Google may end up with more information about patients whose medical information was shared with them in a deidentified fashion than with patients whose medical information was identified in the first place (i.e., with an accompanying business associate agreement limiting reidentification and linkage with other data).

In addition, academic researchers are also increasingly using data and specimens gathered by private industry (thus the ‘revolving door’ analogy). Indeed, the number of peer-reviewed publications using genetic data from a private databank increased from four in 2011 to 57 in 2017 (for a total of 181 over the time period). The vast majority.

---

128 Rebecca Robins, Contract offers unprecedented look at Google deal to obtain patient data from the University of California, Feb. 26, 2020 https://www.statnews.com/2020/02/26/patient-data-contract-google-university-of-california/ (accessed July 16, 2020).

129 Stanford Medicine, Stanford Medicine, Google team up to harness power of data science for health care (Aug. 8, 2016) https://med.stanford.edu/news/all-news/2016/08/stanford-medicine-google-team-up-to-harness-power-of-data-science.html (accessed Feb. 13, 2021) (‘Together, Stanford Medicine and Google will build cloud-based applications for exploring massive health-care data sets, a move that could transform patient care and medical research . . . ‘We are excited to support the creation of the Clinical Genomics Service by connecting our clinical care technologies with Google’s extraordinary capabilities for cloud data storage, analysis and interpretation, enabling Stanford to lead in the field of precision health’, said Pravene Nath, chief information officer . . . ’).

130 Cohen & Mello, Big Data, supra note 70, at 1141 (‘Once an individual’s identity is ascertained, the company could then link EHR data with other types of information about that person (e.g., what they purchase). HIPAA bars none of this except the release of date stamps, and would not be implicated, for example, if Google identified individuals by linking EHR data without HIPAA identifiers to internet data of consumers who visited the University of Chicago hospital and searched online for information about particular medical conditions or if a social-media company linked such EHR data to a user’s posts about a hospital stay.’).

131 Id. at 1084.

132 Id. at 1095.

133 Google and UCSF, Data Evaluation License Agreement, Mar. 1, 2016 https://www.statnews.com/2020/02/26/patient-data-contract-google-university-of-california/ (accessed July 16, 2020).
of those publications (86%) listed at least one academic author.\textsuperscript{134} In fact, the models trained by Google from the UChicago EMR data resulted in a publication in \textit{npj Digital Medicine} with a first and last author from Google and three academic researchers from UChicago (where the data were from) as well as UCSF and Stanford (which have other agreements with Google\textsuperscript{135,136}).

Having data published in the peer-reviewed literature is itself a business asset for industry. Unlike some tangible common resources like fish in a lake, knowledge and ideas are not generally depleted by use.\textsuperscript{137} In fact, the use of knowledge increases its value.\textsuperscript{138} Publishing articles in the peer-reviewed literature also demonstrates the scientific acceptability of a dataset. For example, in the beginning there were serious questions regarding the validity of self-reported phenotypic data, but 23andMe has established such collaborations as a viable research model.\textsuperscript{139} 23andMe publicizes these relationships and publications in order to recruit other research partners.\textsuperscript{140}

We can only speculate regarding why any one researcher might choose to partner with private industry rather than analyze data held by academia, a government entity, or recruit new participants.\textsuperscript{141} But although they may be acting rationally as an individual, many researchers acting in this way can enhance private data resources at the cost of developing and supporting more accessible ones. If researchers continue to or increasingly rely on private datasets over refining, validating, and contributing to other more accessible datasets with their own work—much like Hardin’s overgrazing sheep\textsuperscript{142}—they may unintentionally as a community contribute to the current environment in which accessible banks struggle to compete with the value and size of private sets.\textsuperscript{143} This system contributes to the enabling of private databanks by AMCs either intentionally or unintentionally investing in, and validating, a structure which they do not ultimately control.

\begin{itemize}
\item \textsuperscript{134} Kayte Spector-Bagdady et al., \textit{Genetic Data Partnerships: Academic Publications with Privately Owned or Generated Genetic Data}, 21 GENET MED. 2827, 2828 (2019).
\item \textsuperscript{135} Stanford Medicine, \textit{supra} note 129.
\item \textsuperscript{136} Robins, \textit{supra} note 128.
\item \textsuperscript{137} Brett M. Frischmann et al., \textit{Common Knowledge}, 362 SCIENCE 1240 (2018).
\item \textsuperscript{138} Hess & Ostrom, \textit{supra} note 170.
\item \textsuperscript{139} Catherine Offord, \textit{The Rising Research Profile of 23andMe}, Nov. 30, 2017, \url{https://www.the-scientist.com/news-analysis/the-rising-research-profile-of-23andme-30564} (accessed July 18, 2020) (‘But the value of such self-reported data sets is perceived more highly than it used to be, in part thanks to the success of 23andMe’s research contributions, notes Weinberg. “Historically, people were very skeptical you’d be able to collect data in this relatively simplistic way and still yield the results”, he says. “But I think they have proven again and again that you can do that. There is strength in numbers.”’).
\item \textsuperscript{140} 23andMe, \textit{23andMe for Scientists}, \url{https://research.23andme.com/} (accessed Dec. 29, 2019) (‘The 23andMe cohort is the largest re-contactable research database of genotypic and phenotypic information in the world. By inviting customers to participate in research, we have created a new research model that accelerates genetic discovery and offers the potential to more quickly garner new insights into treatments for disease.’).
\item \textsuperscript{141} But see NIH RePORT, \textit{Genetic data sharing partnerships: Enabling equitable access within academic/private data sharing agreements}, PI: Kayte Spector-Bagdady (‘This research proposes to characterize and evaluate the factors influencing these genetic data partnerships (beginning with academics), compare market drivers to current existing governance structures, and offer a model for best practices.’).
\item \textsuperscript{142} Hardin, \textit{supra} note 26.
\item \textsuperscript{143} John M. Conley et al., \textit{Myriad After Myriad: The Proprietary Data Dilemma}, 15 N. C. J. LAW TECHNOL. 597 (2014).
\end{itemize}
IV. LIMITATIONS IN CURRENT GOVERNANCE OF THE SECONDARY RESEARCH MARKET

In addition to scoping issues, there are several important limitations on the current governance structure of the secondary research market. First, despite extensive and complex informed consent requirements for federally-funded researchers, even when participants provide explicit consent, the result is not consent that is actually informed. Second, despite the recent revision, the regulations written for interventional human research are not tailored adequately for the risk/benefit profiles of secondary research. And third, the combination of the limitations and failures of the human subjects research market has enabled the privatization of this valuable resource—which has related limitations of its own.

IV.A. Breakdown of the Informed Consent Process

A first limitation of the current governance structure is that, even when research consent is obtained, it is often not actually informed.144 Although many have studied how to improve the informed consent process, the only intervention that has been found to consistently improve participant comprehension is a conversation between the prospective participant and a person knowledgeable about the study.145 But the human subjects research regulations, while mandating that many specific things be included in the informed consent form (eg descriptions of risks, benefits, alternatives, confidentiality, compensation, contact information, voluntariness, and information regarding secondary research146) say very little when it comes to the actual informed consent conversation.147

The long lists of mandatory disclosures in research informed consent forms have diminishing returns.148 Participants neither read nor understand them.149 As Meg Leta Jones and Margot Kaminski recently argued:

The U.S. version of individual control and consent is largely understood to be a paper regime, based on long, elaborate privacy policies that nobody reads, and surveillance that is impossible to opt out of in practice. Thus ‘consent’ and ‘notice and choice’ have become somewhat dirty words in data privacy conversations, standing for the exploitation of individuals under the fictional banner of respecting their autonomy.150

144 Ezekiel J. Emanuel & Christine Grady, Case Study. Is Longer Always Better? 38 HASTINGS CENT. REP. 10 (2008) (arguing that consent forms are ‘growing in length and complexity, becoming ever more intimidating, and perhaps inhibiting rather than enhancing participants’ understanding. Participants may not even read them, much less understand them.’).
145 Laura M. Beskow, Lessons from HeLa Cells: The Ethics and Policy of Biospecimens, 17 ANNU. REV. GENOM. HUM. GENET. 409 (2016).
146 45 CFR § 46.166(b) (2019).
147 45 CFR § 46.166(a)(2) (2019) (‘An investigator shall seek informed consent only under circumstances that provide the prospective subject or the legally authorized representative sufficient opportunity to discuss and consider whether or not to participate and that minimize the possibility of coercion or undue influence.’).
148 Patrick Taylor, Personal Genomes: When Consent Gets in the Way, 456 NATURE 32 (2008).
149 Beskow, supra note 145, at 408; Ellen W. Clayton, The Unbearable Requirement of Informed Consent, 19 AM. J. BIOETH. 19 (2019).
150 Meg Leta Jones & Margot E. Kaminski, An American’s Guide to the GDPR 98 DENVER L. REV. 1 (forthcoming) https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3620198 (accessed July 16, 2020).
Governing secondary research use of health data and specimens

In addition, in a recent study of participants who enrolled in a precision medicine trial, the majority of participants had no idea that their data might be commercialized or used for secondary research protocols—mere weeks after they had signed a comprehensive informed consent form disclosing just that.\(^\text{151}\) It appears that the more information the regulations mandate be included in the informed consent form in the name of autonomy and transparency, the less likely participants are to actually read and comprehend it.\(^\text{152}\)

In secondary research—which often involves complex technologies, ethereal risks, and vague protocols—these concerns are compounded.\(^\text{153}\) As Laura Beskow has argued: ‘...informed consent cannot bear the weight it is being asked to shoulder. There is a chasm between the theoretical ideals of informed consent and what it accomplishes in actual practice.’\(^\text{154}\) And Patrick Taylor: ‘We cannot assume that all social goals will be met through a lemming-like coincidence of universal consent.’\(^\text{155}\)

People have a hard time grasping the concept of risk in general.\(^\text{156}\) And, although a primary risk of secondary research is reidentification, the concept of identifiability is itself a false-binary.\(^\text{157}\) The risk of reidentification is on a spectrum of how much risk, which constantly evolves as additional databases are added to public accessibility and emerging aggregation and algorithmic technologies. In addition, people feel differently about the risk of reidentification as it is related to different kinds of data. As Raymond De Vries and Tom Tomlinson argue:

Donors need to know not only whether they can be reidentified; they also need to be able to decide whether the harms caused by reidentification are too high. The answer to that question depends on the type of research findings being protected and the implications for the person's welfare should those findings be disclosed, not just the statistical likelihood of reidentification.\(^\text{158}\)

Even for regulated research with myriad protections regarding informed consent, the laws only practically require documenting legal consent, as opposed to ensuring

---

\(^\text{151}\) Kayte Spector-Bagdady et al., ‘My research is their business, but I’m not their business’: Patient and Clinician Perspectives on Commercialization of Precision Oncology Data 25 THE ONCOLOGIST 620 (2020) (‘Several patient- and clinician-participants did not understand that the consent form already permitted commercialization of patient genetic data and expressed concerns regarding who would profit from the data, how profits would be used, and privacy and access.’).

\(^\text{152}\) Emanuel & Grady, supra note 144.

\(^\text{153}\) Richards & Hartzog, supra note 82, at 1464 (‘Consent is undeniably powerful, and often very attractive. But we have relied upon it too much, and deployed it in ways and in contexts to do more harm than good, and in ways that have masked the effects of largely unchecked (and sometimes unconscionable) power.’).

\(^\text{154}\) Beskow, supra note 145, at 408.

\(^\text{155}\) Taylor, supra note 148, at 32.

\(^\text{156}\) Brian J. Zikmund-Fisher, Helping People Know Whether Measurements Have Good or Bad Implications: Increasing the Evaluability of Health and Science Data Communications, 6 POL’Y INSIGHTS FROM BEHAVIORAL & BRAIN SCIENCES 29, 31 (2019) (‘people tend to ignore single risk statistics in decision making . . .’).

\(^\text{157}\) Kayte Spector-Bagdady et al., Distinguishing Cell Line Research, 5 JAMA ONCOL. 406, 409 (2019) (‘Given technological indeterminacy and individual contributor preference, however, identifiability should not be considered a binary concept. Guidance regarding deidentification under the Privacy Rule of [HIPAA] acknowledges that, even for technically deidentified data, some risk of reidentification will always remain. How much risk is acceptable is ultimately a policy question that requires ethical analysis.’).

\(^\text{158}\) Tom Tomlinson & Raymond G. De Vries, Human Biospecimens Come from People, 41 ETHICS HUM. RES. 22, 23 (2019).
Governing secondary research use of health data and specimens

consent that is actually informed. And that necessity to document can also lead to limiting research with important populations who could otherwise benefit from it, like cognitively impaired adults.159 As Taylor has bemoaned: ‘Ethics is reduced to autonomy; autonomy is reduced to naked choice; and a self-commodifying model of choice is substituted for richer visions of human nature and interdependence.’160

IV.B. Regulations are not Responsive to Current Secondary Research Market

A second problem with the current regulatory structure surrounding research data and biospecimens is that it was a system originally built to protect people from potentially harmful interventions, not secondary research. The regulations are therefore not calibrated to the unique risk–benefit profiles of secondary research, leading to both over- and under-regulation. Also, the regulations’ threshold for governance is still method of collection—which aligns with neither contributor concerns regarding usage nor current data-sharing practices.

IV.B.1. Regulations Lack Responsiveness to the Risk/Benefit Profiles of Secondary Research

A major issue with the current governance of the federally funded secondary research enterprise is a lack of responsiveness to the risk/benefit profiles of most secondary research protocols. Not only does secondary research need large amounts of biospecimens and data in the aggregate to be helpful—distilling the value of any one participant—lowered risks to individuals, level of burden if those risks materialize, and greater benefits to the community warrant a reconsideration of the current informed consent requirements.161

With the transition from interventional clinical research to secondary research protocols, corresponding risks to participants have also shifted.162 Whereas the research violations that founded the current regulatory and case law structure were tangible and largely physical163 or financial164 in nature, secondary research risks generally involve dignitary harms which are harder to quantify165 and establish as damages in court.166 Participants in secondary research protocols can suffer from violations to what some have dubbed ‘non-welfare interests’, or the ‘moral, religious, or cultural concerns’ about

159 Beth Prusaczyk et al. Informed Consent to Research with Cognitively Impaired Adults: Transdisciplinary Challenges and Opportunities, 40 CLIN. GERONTOL. 1, 63–73 (2017).
160 Taylor, supra note 148, at 32.
161 K. Spector-Bagdady & J. Beever, Rethinking the Importance of the Individual within a Community of Data HASTINGS CENT REP doi: 10.1002/hast.1112 (2020) (‘Given the low-individual-risk and high-community-benefit profile of many secondary research protocols, in which biospecimens generally play a large role, to subjugate (high) community benefit to (low) individual risk at this scale is inappropriate. We should instead be prioritizing assessment of risk versus benefits at the community level.’).
162 Rai, supra note 1, at 1651 (‘... 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.’).
163 Jones, supra note 30.
164 Moore, 51 3d at 146.
165 Price & Cohen, supra note 103, at 40 (‘Especially for deontological concerns with health privacy, the loss of control over who accesses an individual’s data and for what purpose matters, even if there are no material consequences for the individual or if the individual does not even know.’).
166 Parks v. Wells Fargo Home Mortg., Inc., 398 F.3d 937, 940–41 (7th Cir. 2005) (quoting Maere v. Churchill, 116 Ill. App. 3d 939, 944, 452 N.E.2d 694, 697 (3d Dist. 1983)). Dinerstein, 2020WL 5296920.
uses of their data or specimens, even when they are never reidentified (also related to the concept of a ‘dignitary tort’ in law, such as intentional infliction of emotional distress). For example, in one study by De Vries and colleagues, whereas more than 70% of participants surveyed said they would be happy to sign a ‘blanket consent’ for any future use of their donated biospecimen, when pressed specifically about controversial examples of research—such as those involving patents, abortions, or weapons of mass destruction—almost as many changed their mind and asked to withdraw.

Last, the value of the denominator in research has shifted considerably. In interventional clinical research, due to the intensive control of variants and attempts to ensure that the study is fully powered to capture statistically significant differences, each participant can be of value to the study at the individual level. But in secondary research protocols participants are generally of value in aggregate. This is analogous conceptually to the recognized ‘prevention paradox’ in public health, where changes in health behavior can positively affect outcomes at the population level but only result in negligible improvements for any given individual.

On the bright side, due to this shift in value—and unlike, for example, the single cancer patient who may be enrolled in only one experimental chemotherapeutic or standard of care control arm—uses of aggregate data are generally ‘non-rivalrous’ or ‘non-subtractive’ of others’ uses. As Charlotte Hess and Elinor Ostrom, the founders of modern common resource policy, have argued: ‘...the more people who share useful knowledge, the greater the common good.’

Last, upon reviewing the new types of risks and benefits for secondary research, an important overarching observation becomes evident. Although the risks of secondary research (eg privacy breaches, dignitary harms) remain remote, small, and individual, the benefits (valuable datasets and knowledge outputs) redound to either the entity holding the data or the common good. This situation is similar to the development of ‘herd immunity’ in the public health context, where for some diseases 90% of individuals must be vaccinated to achieve herd immunity of a community that otherwise protects those who cannot be vaccinated. The pursuit of herd immunity in vaccinations relies on social contract theory and requires individuals to take on some small risk to
themselves (of adverse events associated with the vaccine) to benefit others greatly (by protecting them from the more harmful disease).

Similarly, the asymmetry between the risk- and benefit-bearer can have detrimental effects on an efficient secondary research biospecimen and data market. Market actors often make decisions based on costs and anticipated benefits for themselves, and disregard or discount costs to others. This ‘negative externality’,\textsuperscript{174} of researchers discounting privacy risks to the participants of secondary research protocols, can cause the secondary research market to become inefficient, ie not fully tally the actual costs and benefits of an action-unless externally forced to perform otherwise.

\textbf{IV.B.2. Governance by Methods of Collection No Longer Makes Sense}

Another problem with the current governance of the secondary research enterprise is that siloed regulation of data and biospecimens by collection entity no longer makes sense due to increased sharing. This is why the new data regulations in both California and Europe generally regulate by the \textit{kind} of data being used, rather than by \textit{who} is using it.\textsuperscript{175} Also, in the USA, who collected the data is an irrelevant distinction to contributors who generally care about its use.

There has been a plethora of empirical studies demonstrating that people care how their data and specimens are used for research.\textsuperscript{176} A minority even worry about research with deidentified data and specimens, which is particularly true in Black and Latino populations,\textsuperscript{177} and currently out of scope for all regulatory regimes. In addition, although many people hypothetically support the use of specimens in research,\textsuperscript{178} biobanks are increasingly turning to a ‘commercialization’ model in order to support the expenses of long-term cryopreservation.\textsuperscript{179} But a recent US-based study found that 67% of participants wanted to be clearly notified regarding potential biospecimen commercialization and only 23% were comfortable with such use.\textsuperscript{180} The revisions to the Common Rule now require that regulated researchers disclose potential commercialization to participants, putting them at a potential additional disadvantage for recruitment as compared to their unregulated peers.\textsuperscript{181} As aptly summarized by the recent \textit{National Academy of Medicine} (NAM) report ’Health Data Sharing to Support Better Outcomes: Building a Foundation of Stakeholder Trust’:

\begin{verbatim}
174 Fundamental Finance, Negative Externality, http://economics.fundamentalfinance.com/negative-externality.php (accessed April 6, 2020).
175 Price et al., Shadow Health Records, supra note 94, at 450.
176 Spector-Bagdady, Encouraging Participation, supra note 8.
177 Jagsi et al., supra note 9.
178 Jeffrey Peppercorn et al., Patient Preferences for Use of Archived Biospecimens from Oncology Trials When Adequacy of Informed Consent Is Unclear, 25 The Oncologist 78 (2019).
179 Timothy Caulfield et al., A Review of the Key Issues Associated With the Commercialization of Biobanks, 1 J. Law Biosci. 94 (2014).
180 Spector-Bagdady, Encouraging Participation, supra note 8.
181 Jessica L. Roberts, Negotiating Commercial Interests in Biospecimens, 45 J. Law Med. Ethics 138 (2017); Joshua D. Smith et al., Immortal Life of the Common Rule: Ethics, Consent, and the Future of Cancer Research, 35 J. Clin. Oncol. 1879 (2017).
\end{verbatim}
The patient and family community lacks trust that health care systems and researchers will make data and the conclusions based on those data available to them and will not misuse data they provide by rationing care and sharing it with unauthorized third parties.\textsuperscript{182}

And last, even if the regulations are applicable and consent is obtained, it does not provide the kind of control that contributors want. The Common Rule or HIPAA only provide a binary ‘exit right’—the right to either contribute to research or not.\textsuperscript{183} Participants are not given a voice in the kind of research gets done or veto power over what secondary research protocols they are or are not willing to contribute.\textsuperscript{184} As Richards and Hartzog argue: ‘...consent does not scale. It is almost entirely incompatible with the modern realities of data and technology in all but the most limited of circumstances.’\textsuperscript{185}

\textbf{IV.C. Encouraging Privatization of a Shared Resource}

A third major problem with the current governance of the secondary research market is that it actually enables the privatization of secondary research data and biobanks. And privately-held databanks, as opposed to those run by a government agency or other type of accessible collaboratory, are associated with several pressing societal and scientific concerns.

\textbf{IV.C.1. Privatization of Data and Biobanks}

As discussed above, the proposed alternative to governance of shared resources in the ‘tragedy of the commons’, other than government regulation, is privatization. As Michael Heller and Rebecca Eisenberg pointed out (as far back as 1998), this is the direction biomedical research is headed.\textsuperscript{186} But it is not fair to assume that the privatization of a shared resource is necessarily bad. According to Hardin, it is the only other option in averting sure ruin.\textsuperscript{187}

The US federal and global governments have focused vast resources into encouraging data accessibility and sharing. The European Open Science Cloud and various country-specific genetic and other health data and specimen banks have aggressively

\textsuperscript{182} NAM, \textit{Health Data Sharing to Support Better Outcomes: Building a Foundation of Stakeholder Trust} (2020) https://nam.edu/health-data-sharing-special-publication/ (accessed Jan. 28, 2021).

\textsuperscript{183} Evans, supra note 4, at 684 (‘These federal regulations conceive individual protection as an exit right (informed consent/authorization) while granting people no real voice in setting the goals of informational research or the privacy and ethical protections they expect. Lacking a voice, people CRexit from research—that is, they exercise their Common Rule, or “CR”, informed consent rights to exit from informational research altogether. This CRexit strategy scarcely advances people’s interests, when surveys show that most Americans would like to see their data used to advance science.’).

\textsuperscript{184} Id. (‘Low participation in informational research should be viewed as what it is: a widespread popular rejection of the unsatisfactory, top-down protections afforded by existing regulations.’).

\textsuperscript{185} Richards & Hartzog, supra note 82, at 1467.

\textsuperscript{186} Michael A. Heller & Rebecca S. Eisenberg, \textit{Can patents deter innovation? The anticommons in biomedical research}, 280 \textit{SCIENCE} 698, 698 (1998) (‘Since Hardin’s article appeared, biomedical research has been moving from a commons model toward a privatization model... Today, upstream research in the biomedical sciences is increasingly likely to be “private” in one or more senses of the term—supported by private funds, carried out in a private institution, or privately appropriated through patents, trade secrecy, or agreements that restrict the use of materials and data.’); see also Craig Konnoth, Preemption through Privatization, \textit{Harvard L. Rev}, Vol 134 (2021).

\textsuperscript{187} Hardin, supra note 26, at 1244.
pursued the ideal of open science abroad. The NIH’s 2018 Strategic Plan for Data Science proposes a ‘data ecosystem’ which would allow for ‘a distributed, adaptive, open system with properties of self-organization, scalability and sustainability’ via projects including the NIH Data Commons. However, these initiatives have yet to achieve widely engaged data-sharing practices.

A decade ago, the NAM report Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease argued:

Data-sharing standards should be created that respect individual privacy concerns while enhancing the deposition of data into the Information Commons. Importantly, these standards should provide incentives that motivate data sharing over the establishment of proprietary databases for commercial intent. Resolving these impediments may require legislation and perhaps evolution in the public’s expectations with regard to access and privacy of health-care data.

This report, in turn, inspired the US government’s dedication to building a health biospecimen and data commons, the All of Us Research Program, which was announced by President Barack Obama in his 2015 State of the Union address. The goal of All of Us is to ‘enroll at least 1 million persons who agree to share their EMR data, donate biospecimens for genomic and other laboratory assessments, respond to surveys, and have standardized physical measurements taken.’ It supports ~270,000 participants, and recently started returning a first round of genic results to participants recruited by the University of Wisconsin. All of Us targets recruitment to those historically ‘underrepresented in biomedical research’, and it currently boasts 75% of participants that meet that categorization. Since 2015, Congress has allocated $1.02 billion toward supporting the All of Us program and the 21st Century Cures Act authorized another $1.14 billion through 2026.

By contrast, the DTC-PGT company 23andMe has a genetic database of over 12 million participants and counting, as well as over a billion phenotypic data points. This makes it ‘the largest re-contactable research database of genotypic and phenotypic information in the world.’ And 23andMe recently announced plans to go public with

188 Mark Phillips & Bartha M. Knoppers, Whose Commons? Data Protection as a Legal Limit of Open Science, 47 J. L.A.W.M. 106 (2019).
189 NIH, Strategic Plan for Data Science, 2018, https://datascience.nih.gov/sites/default/files/NIH_Strategic_Plan_for_Data_Science_Final_508.pdf (accessed July 18, 2020).
190 Ida Sim et al., Time for NIH to Lead on Data Sharing, 367 SCIENCE 1308 (2020).
191 National Research Council, supra note 3.
192 All of Us Research Program Investigators, The ‘All of Us’ Research Program, 381 N. ENGL. J. MED. 668, 668 (2019).
193 All of Us Research Program, National Research Program Returns First Results to Local Participants (Jan. 21, 2021) https://allofus.nih.gov/news-events-and-media/news/national-research-program-results-first-results-local-participants (accessed Feb. 13, 2021).
194 Id. at 669 (‘Among persons from whom biospecimens are obtained, the target percentage of persons in racial and ethnic minorities is more than 45% and that of persons in underrepresented populations is more than 75%.’)
195 Id. at 675.
196 23andMe, About Us, https://mediacenter.23andme.com/company/about-us/ (accessed Aug. 15, 2019).
197 23andMe, 23andMe therapeutics, https://therapeutics.23andme.com/ (last visited July 17, 2020).
a merger valuation of $3.5 billion.\footnote{Kari Paul, Fears over DNA privacy as 23andMe plans to go public in deal with Richard Branson, The Guardian (Feb. 9, 2021) https://www.theguardian.com/technology/2021/feb/09/23andme-dna-privacy-richard-branson-genetics (accessed March 16, 2021).} Therefore, despite All of Us’ laudable goals and progress, the US government has essentially spent $2.16 billion to build a database to compete with the private ones its own regulations enabled in the first place.

\section*{IV.C.2. Challenges with Privatization}

There are costs associated to allowing a valuable common resource to become privatized. First, informed consent is lacking at an even greater scale. Second, when data and biospecimens are privately held, industry can put limitations on access that might stifle future research advances. Third, without data access, peer researchers can neither validate nor build on work otherwise available in the peer-reviewed literature. Last, even if certain industry players are currently allowing limited researcher access to their datasets, a new business focus, leadership, or regulation could change that arrangement quickly—and the time and effort other researchers spent contributing to and validating that private resource could be lost.

First, issues with the Common Rule’s informed consent process can be compounded in unregulated industry research which generally relies on digital consent platforms (if it acquires consent at all). This type of consent to private consumer interactions is not a typical clinical or research informed consent with associated fiduciary obligations—it is contractual.\footnote{Kaye Spector-Bagdady, Reconceptualizing Consent for Direct-to-Consumer Health Services, 41 Am. J. Law Med. 568 (2015).} There is generally no IRB ensuring that the risks and benefits are adequately balanced before attempting to enroll participants, but rather an attorney (best-case scenario) who has wordsmthed language to protect her company from liability and grant them latitude. Richards and Hartzog highlight three conditions that might make electronic consent particularly ‘pathological’: (i) when people are asked constantly for their consent such that there is not enough time to seriously consider each choice, (ii) when the risks of consent are complex and ethereal, and (iii) when people are incentivized to not take choices seriously\footnote{Richards & Hartzog, supra note 82, at 1461 (‘We argue that consent is most valid when we are asked to choose infrequently, when the potential harms that result from the consent are easy to imagine, and when we have the correct incentives to consent consciously and seriously. The further we fall from this gold standard, the more a particular consent is pathological and thus suspect.’).}—all of which are relevant to the e-consent of unregulated research.

People also generally do not think about inputting their health data into private platforms, such as ‘wellness apps’ or industry websites, as commercializing their own information.\footnote{Kirsten Ostherr et al., Trust and privacy in the context of user-generated health data, Big Data & Society 1 (2017) (‘Members of the general public expressed little concern about sharing health data with the companies that sold the devices or apps they used...’).} Almost a quarter of US adults report that they are asked to agree to a private privacy policy on a near daily basis.\footnote{Brooke Auxier et al., Americans and Privacy: Concerned, Confused and Feeling Lack of Control Over Their Personal Information, Nov. 15, 2019, https://www.pewresearch.org/internet/2019/11/15/americans-and-privacy-concerned-confused-and-feeling-lack-of-control-over-their-personal-information/ (accessed Nov. 15, 2019).} Only 1/1000 consumers click on a website’s terms of service; only 1/10,000 if it requires two clicks. For the very few
who make it to terms of service, the median time spent reading them is 29 seconds.\textsuperscript{203} Contemplating sharing sensitive health data does not change this.\textsuperscript{204} And, even if consumers do glance at the terms and conditions, the risks of secondary usages of data are so complex that there is disagreement regarding whether terms and conditions governing them should even be considered valid.\textsuperscript{205} This also explains contributor concern in the summer of 2019 when 23andMe announced its exclusivity agreement with GlaxoSmithKline to use its database for drug research and development (despite the fact that such a potential collaboration was laid forth clearly in 23andMe’s Terms and Conditions).\textsuperscript{206}

In addition to the fact that most people do not read informed consent forms—and even if they do, comprehension is likely fleeting—people are also generally unaware of the possibility of ‘data mosaicking’ (combining different datasets to gain a more complete picture of a single point of interest):

Our consent to data practices is astonishingly dispersed. Thousands of apps and services ask us for small, incremental disclosures, few of which involve the kind of collection of information that might give people pause. While dating apps and platforms that collect sensitive and large amounts of personal data might cause some pause, it’s not as though people share all their information at once. Instead, it trickles out over time, such that people’s incentives to deliberate at the point of agreement are small because we don’t know how much information we will ultimately end up sharing.\textsuperscript{207}

Indeed, companies that create ‘shadow health records’ have started to multiply and gather health-related and proxy data from nonprotected sources, assemble the data back together under the identity of the individual from whom it came, and then sell access to the Frankenstein-ed health records back to researchers.\textsuperscript{208}

And, although the benefits of digital consent are generally immediate and obvious (eg a funny picture of you as the opposite gender, or 30 years older, to share on social media\textsuperscript{209}), deidentification and mosaicking risks are often unknown, complex to grasp,
and may or may not become actual burdens.\textsuperscript{210} Few understand that researchers will not just have the ability to access to your ‘selfie’, but connect it with troves of other data you share—indeed indefinitely. A recent example of this is predators collecting name, birthday, and location information from people posting photos of themselves with their COVID-19 vaccination cards for potential scams or the sale of fake cards.\textsuperscript{211}

A second issue is potential limits on access. Researcher use of private datasets allows industry a gatekeeping function over what research is enabled. 23andMe has an internal committee that reviews proposed protocols and only supports a select few per year.\textsuperscript{212} But gatekeeping can compromise the reliability of peer-reviewed literature and bias it in favor of industry. In 2011, for instance, 96.5\% of published, industry-sponsored, head-to-head comparative effectiveness trials found favorable results—a highly unlikely outcome, presumably associated with the type of reports submitted for publication.\textsuperscript{213} These kinds of access issues are already well-documented in the medical drug market.\textsuperscript{214} But, as biospecimens and health data emerge as increasingly critical to drug and device development,\textsuperscript{215} it is reasonable to anticipate similar problems in the upstream data market.

Third, if datasets are private, other researchers’ ability to validate the work or build derivative discoveries will be limited. For example, there has been a recent rash of reanalyses of studies in the nutritional\textsuperscript{216} and psychological literature\textsuperscript{217} which has debunked many major studies—a critical re-examination opportunity that supports the integrity of science in the published literature. As argued in the recent NIH \textit{Policy for Data Management and Sharing}:

\begin{quote}
Sharing scientific data accelerates biomedical research discovery, in part, by enabling validation of research results, providing accessibility to high-value datasets, and promoting data reuse for future research studies.\textsuperscript{218}
\end{quote}

\textsuperscript{210} Pike, supra note 22, at 710 (‘Finally, the reality is that people are imperfect decision makers, particularly with choices that involve immediate gratification and delayed, but uncertain, consequences.’); Richards & Hartzog, supra note 82, at 1497 (‘people would have little incentive to deliberate because, frankly, they have little notion of the stakes, and the benefits of consent are right at their fingertips.’).

\textsuperscript{211} Better Business Bureau, BBB Tip: Do not share your COVID-19 vaccine card on social media (Jan. 29, 2021) https://www.bbb.org/article/news-releases/23675-bbb-tip-dont-share-your-vaccine-card-on-social-media (accessed Nov. 13, 2021).

\textsuperscript{212} 23andMe, 23andMe Research Innovations Collaborations Program https://research.23andme.com/research-innovation-collaborations/ (accessed Aug. 22, 2019) (‘We accept applications from academic researchers on a rolling basis. In June and December, we hold a scientific review to evaluate proposals for the limited number of collaborative projects we can initiate. Applicants are informed of our committee decision approximately three months after the deadline.’).

\textsuperscript{213} Maria Elena Flacco et al., Head-to-Head Randomized Trials are Mostly Industry Sponsored and Almost Always Favor the Industry Sponsor, 68 J. CLIN. EPIDEMIOL. 811 (2015).

\textsuperscript{214} Eg C. Lee Ventola, The drug shortage crisis in the United States: Causes, Impact, and Management Strategies, 36 PT. 740 (2011); and Daniel Kozarich, Mylan’s EpiPen Pricing Crossed Ethical Boundaries, FORTUNE, Sept. 27, 2016, http://fortune.com/2016/09/27/mylan-epipen-heather-bresch/ (accessed Apr. 2, 2018).

\textsuperscript{215} Brodwin & Palmer, supra note 125.

\textsuperscript{216} John P. A. Ioannidis, The Challenge of Reforming Nutritional Epidemiologic Research, 320 JAMA 969 (2018).

\textsuperscript{217} Adam Marcus, Psychological Science in the news again: CNN retracts story on hormone-voting, Oct. 25, 2012, https://retractionwatch.com/2012/10/25/psychological-science-in-the-news-again-cnn-retracts-story-on-hormone-voting-link/ (accessed July 18, 2020).

\textsuperscript{218} NIH, Policy for Data Management, supra note 2.
In fact, two COVID-19 drug studies from the same authorship team were recently retracted from the *New England Journal of Medicine*[^19] and *The Lancet*[^20], respectively due to a lack of author-access to privately-held data—a fact which authors misrepresented when submitting the articles and confirming data validity.

Without access to underlying data sources, this type of post-hoc quality assurance is much harder, if not impossible.[^21] And, as Heller and Eisenberg argued, too many intellectual property rights in premarket ‘upstream’ research can result in limited ‘downstream’ produce-oriented research in the ‘tragedy of the anti-commons’. This type of tragedy can, in turn, lead to underutilization of a valuable resource.[^22]

Although norms are beginning to shift gradually, extensive disclosure of raw data in supplements to publications remains uncommon. If an external researcher were later to request underlying data from the corresponding author for verification or further research, the response to such query would also generally not be public (and there are many reports of data actually being notably unavailable upon request).[^23] This tension is exacerbated when the underlying dataset has commercial value.[^24] But, even if databanks are not protected by patents or other types of business interests, and even if corresponding authors would be willing to share the underlying datasets supporting their work, the burden of requesting and accessing the dataset is still on other researchers:

... when it is easy for owners to exclude users from access to resources, as in the case of ‘practically excludable’ materials and data, the burden of inertia is on users to persuade owners to permit access, whether or not the resource is covered by formal property rights

[^19]: Mandeep R. Mehra et al., *Retraction: Cardiovascular Disease, Drug Therapy, and Mortality in Covid-19*, N. Engl. J. Med. DOI: 10.1056/NEJMoa2007621, 382 N. Engl. J. Med. 26, at 2582 (2020) (‘Because all the authors were not granted access to the raw data and the raw data could not be made available to a third-party auditor, we are unable to validate the primary data sources underlying our article... We therefore request that the article be retracted.’)

[^20]: Mandeep R. Mehra et al., *RETRACTED: Hydroxychloroquine or Chloroquine With or Without a Macrolide for Treatment of COVID-19: A Multinational Registry Analysis*, LANCET doi: 10.1016/S0140-6736(20)31180-6 (2020).

[^21]: Rebecca S. Eisenberg, *Proprietary Rights and the Norms of Science in Biotechnology Research*, 97 Yale L.J. 177, 197 (1987) (‘But for research involving the use of unique biological materials, such as bacterial strains and other types of self-replicating cells, publication in writing alone may not be sufficient to satisfy this replicability norm. To replicate the authors’ results, subsequent investigators may need access to identical materials. By sharing access to unique materials, however, the publishing scientist not only enables other scientists to replicate her claims; she also allows them to compete with her more effectively in making new discoveries.’) (hereinafter, Eisenberg, *Proprietary Rights*).

[^22]: Heller & Eisenberg, supra note 185, at 698 (‘A resource is prone to overuse in a tragedy of the commons when too many owners each have a privilege to use a given resource and no one has a right to exclude another. By contrast, a resource is prone to underuse in a “tragedy of the anticommons” when multiple owners each have a right to exclude others from a scarce resource and no one has an effective privilege of use.’).

[^23]: Eisenberg, *Proprietary Rights*, supra note 220, at 198 (‘Withholding materials is a relatively inconspicuous departure from scientific norms. It occurs after publication and is not apparent from the written text.’).

[^24]: Id. (‘... publishing scientists with exclusive access to such materials have an opportunity to gain recognition while retaining a future advantage over their research competitors. This conflict between norms and incentives is aggravated when the materials (or the discoveries they facilitate) have potential commercial value.’).
such as patents. In this context, high transaction costs make use less likely, aggravating the risk of an anticommons.\textsuperscript{225}

Last, companies might decrease or limit researcher access because of a new business focus, leadership, or regulation. For example, when law enforcement used the GEDmatch database surreptitiously to solve the ‘Golden State Killer’ case, there was substantial backlash.\textsuperscript{226} GEDmatch only listlessly pursued claims against the law enforcement officials who violated their terms and conditions by making up an identity.\textsuperscript{227} The company was then bought by Verogen, a forensic genomics firm, whose specific intent was to aid in law enforcement searches. The database, including the genomic data of over 1.3 million users, was sold \textit{en masse} without so much as an affirmative notice to contributors. GEDmatch then updated its Terms of Service to disclose the acquisition, and users were locked out of accessing the platform until they agreed to the new terms.\textsuperscript{228} Relatedly, in February 2021, 23andMe announced its decision to go public—a move for which the implications are not yet clear.\textsuperscript{229}

V. AMCS SHOULD CREATE NEW POLICIES TO CONTROL THE REVOLVING DOOR BETWEEN ACADEMIA AND INDUSTRY

We know that what contributors care about is use of their health data and specimens.\textsuperscript{230} We know that they are particularly suspect when health data and specimens exchange hands across types of entities.\textsuperscript{231} And we know the government is currently regulating the secondary research enterprise in a way that remains nonresponsive to either of those concerns. But questions remain: Should someone do something about it? And, if so, who?

There have been several sets of important recommendations regarding better governing biospecimens and health data research, which generally segregate along four main suggestions: strengthen existing or pass new law, require all users of the system to contribute, develop accessible resources, or enable data commons.\textsuperscript{232} Many

\begin{footnotesize}
\begin{itemize}
\item \textsuperscript{225} Rebecca S. Eisenberg, \textit{Noncompliance, Nonenforcement, Nonproblem? Rethinking the Anticommons in Biomedical Research}, 45 Hous. L. Rev. 1059, 1098–99 (2008) (hereinafter, Eisenberg, \textit{Noncompliance}).
\item \textsuperscript{226} Avi Selk, \textit{The ingenious and 'dystopian' DNA technique police used to hunt the 'Golden State Killer' suspect}, NY Times, Apr. 28, 2018, https://www.washingtonpost.com/news/true-crime/wp/2018/04/27/golden-state-killer-dna-website-gedmatch-was-used-to-identify-joseph-deangelo-as-suspect-police-say/ (accessed Dec. 25, 2019).
\item \textsuperscript{227} Cassie Martin, \textit{Why a warrant to search GEDmatch’s genetic data has sparked privacy concerns}, ScienceNews, Nov. 12, 2019, https://www.sciencenews.org/article/why-warrant-search-gedmatch-genetic-data-has-sparked-privacy-concerns (accessed Dec. 25, 2019).
\item \textsuperscript{228} Megan Molteni, \textit{A DNA firm that caters to police just bought a genealogy site}, Wired, Dec. 9, 2019, https://www.wired.com/story/a-dna-firm-that-caters-to-police-just-bought-a-genealogy-site/ (accessed Dec. 25, 2019).
\item \textsuperscript{229} Brodwin & Palmer, supra note 125.
\item \textsuperscript{230} De Vries, supra note 166.
\item \textsuperscript{231} Caulfield, supra note 178.
\item \textsuperscript{232} Cambon-Thomsen, supra note 17; Ruth Chadwick & Kare Berg, \textit{Solidarity and Equity: New Ethical Frameworks for Genetic Databases}, 2 Nat. Rev. Genet. 318 (2001); Robert Cook-Deegan et al., \textit{Introduction: Sharing Data in a Medical Information Commons}, 47 J. Law Med. Ethics 7 (2019); Patricia A. Devorka et al., \textit{Hopeful and Concerned: Public Input on Building a Trustworthy Medical Information Commons}, 47 J. Law Med. Ethics 70 (2019); Evans, supra note 4; Frischmann et al., supra note 137; Hess & Ostrom, supra note 170; Sharona Hoffman, \textit{Electronic Health Records and Medical Big Data: Law and Policy}
\end{itemize}
\end{footnotesize}
scholars have theoretically supported an opt-out or no consent system for public health purposes, although—when asked—few potential contributors agree. But, because many previous proposals either required private industry to self-restrain when competitors might not, or were founded on the government issuing fluid and comprehensive regulatory revisions when it apparently cannot, none have yet fully stemmed the tide of privatization.

This section first explores industry self-governance as one alternative to the current regulatory structure. Given its limitations, it will then discuss alternative approaches AMCs can take to control the commercialization of held health data and biospecimens as well as academic researcher use of industry data, in addition to improving standards for informed consent for both patients and participants.

V.A. Self-Governance as an Alternative?

Stephen Maurer, in his book on Self-Governance in Science, argues that self-governance of industry can be a viable alternative when regulatory mechanisms have proven ineffective. There are many types of industries where self-governance can be attractive, both to ward off potentially more restrictive government regulation, or because the consumer base insists on a standard for behavior without assurance of which they will no longer purchase the product. In addition, private industry players are in a position to possess the most relevant information regarding effective strategies for controlling industry behavior. They may also be able to reach a larger swath of players than government agencies, which are constrained by congressional scope and funding.

In particular, industries where players are limited and derivations from market expectations may be attributed across all entities might be particularly motivated to self-regulate. For example, when 23andMe sales started to decline in 2018, CEO

(2016); Pike, supra note 22; Rai, supra note 1; Richards & Hartzog, supra note 82; Mark D. Wilkinson et al., The FAIR Guiding Principles for scientific data management and stewardship, 15 Sci. DATA. 160018 (2016).

233 Amy L. McGuire, supra note 24, at 15 (2019) (‘We acknowledge that weighing all relevant considerations, opt-out or no consent may be appropriate in contexts such as public health surveillance, use of de-identified information from a single source, like an electronic health record or newborn blood spot collection, or for research where an inclusive dataset is genuinely necessary in order to produce unbiased results, subject to careful oversight and accountability mechanisms.’).

234 Juli M. Bollinger et al., What is a Medical Information Commons? 47 J. L. MED. & ETHICS. 41, 46 (2019). (‘There was little enthusiasm for MICs to operate under a public health model where individual participation is mandatory or there is an opt-out consent option rather than opt-in. Interviewees commented that an opt-out model would require a high level of trust that is unlikely to be found in the U.S. population given past incidents of research misconduct (and concerns about discrimination in healthcare and other domains), especially among minority/underrepresented populations.’).

235 Stephen M. Maurer, Self-Governance in Science, at 180 (2017) (‘conventional government has tried and failed, sometimes repeatedly, to address particular policy problems.’).

236 Id., at 239 (‘Nearly all scholars agree that firms are more likely to organize self-governance when downstream markets are “highly branded” so that demand depends less on objective attributes than “marketing” so that demand depends less on objective attributes than “marketing” and a “constructed brand identity” which makes them vulnerable to “shifts in consumer preferences.”’).

237 Id., at 179 (‘US officials have long recognized that private organizations often possess more information and can produce better standards.’).

238 Id., at 179, 181.

239 Id., at 4 (‘But [the virtuous circle of self-regulation] can also work in reverse, so that each new defector who leaves a standard devalues compliance for those who remain and tempts them to leave as well. The case is
Anne Wojcicki attributed it to privacy concerns surrounding use of their databank by law enforcement after the GEDmatch saga. Sales of the 23andMe product have continued to decline (by $136 M just from 2019 to 20), and last year 23andMe laid off 14% of its employees. Wojcicki attempted to head the generalization of GEDMatch privacy laxness off by coauthoring a new Privacy Best Practices for Consumer Genetic Testing Services within months of the GEDmatch news and in the same week she announced the 23andMe deal with GlaxoSmithKline. 23andMe was joined on these Best Practices by several other major DTC-PGT players including Ancestry and Helix.

Self-regulation also has the potential to be more efficient and finely tuned to market variance than external regulatory bodies. Evolving technology moves faster than notice and comment rulemaking allows. Evocative of when FDA finally intervened with the DTC-PGT market and all other companies but for 23andMe dropped off the market or began requiring a prescription, Maurer points out:

... official regulation can be oversupplied so that politicians and bureaucrats invest more than what the entire industry is worth to society. By comparison, society’s investment in private regulation can never exceed what consumers are willing to pay for the regulated product.

That is, if industry spends more money on self-regulation that it can regain in sales, an efficient market would right itself by expelling such a product (23andMe was able to avoid this market consequence via its investment by Google). In addition, as argued by the recent NAM report on health data sharing:

Standards of conduct can build trust, because people know what to expect. Collaborative efforts built on trust can convert zero-sum relationships into positive-sum relationships, where data sharing serves everyone’s interests

very different where the standard’s benefits are externalities, ie flow indiscriminately to every member of the community. The most important example is where backlash damages firms that had nothing to do with the scandal.’

240 Brodwin & Palmer, supra note 125.
241 Future of Privacy Forum, Privacy Best Practices for Consumer Genetic Testing Services (July 31, 2018), https://fpf.org/wp-content/uploads/2018/07/Privacy-Best-Practices-for-Consumer-Genetic-Testing-Services-FINAL.pdf (‘The Best Practices provide a policy framework for the collection, retention, sharing, and use of Genetic Data generated by consumer genetic and personal genomic testing services.’).
242 GlaxoSmithKline, supra note 205.
243 Future of Privacy Forum, supra note 240 (‘The Best Practices provide a policy framework for the collection, retention, sharing, and use of Genetic Data generated by consumer genetic and personal genomic testing services.’).
244 Wendy E. Parmet, Employers’ Vaccine Mandates Are Representative of America’s Failed Approach to Public Health, ATLANTIC (Feb. 4, 2021) https://www.theatlantic.com/ideas/archive/2021/02/privatization-public-health/617918/ (‘Indeed, the private sector is often seen as nimbler than the government precisely because it can eschew the necessity of public input and the threat of public accountability.’).
245 Maurer, supra note 234, at 156.
246 NAM, supra note 181.
There are some successful examples of health industry self-governance. The Pharmaceutical Research and Manufacturers of America’s (PhRMA) *Code on Interactions with Health Care Professionals* brings additional clarity and specificity to existing FDA regulations and the antikickback law regarding incentives from pharmaceutical sales representatives to health care professionals (e.g., specifying that acceptable ‘gifts’ must be educational in nature). In return, the government has acted creatively to incorporate PhRMA’s perspective and flexibility into governance by stating that compliance may protect companies from liability or even requiring PhRMA Code compliance in settlement or corporate integrity agreements.

However, the private data and biospecimen industry has yet to attempt comprehensive self-regulation. The closest example is the 23andMe-driven DTC-PGT industry’s recent *Privacy Best Practices for Consumer Genetic Testing Services* discussed above. But, most DTC-PGT companies did not actually sign it, and the statement’s scope had notable limitations including a lack of privacy protections for nongenetic health data also collected.

Thus, although self-governance is a potentially effective possibility for the private data and biospecimen industry, it has not yet come to fruition despite decades of concern. Also, as Maurer concludes, ‘this bargain is only available provided government can trust the private process’, and there is no indication that is true.

### V.B. Proposed AMC Policies Regarding Commercialization and Commercial Use of Data

The broad limitations in scope and application of the federal human subjects research regulations put federally funded researchers at a competitive disadvantage vis-à-vis private industry, so much so that AMCs are increasingly buying their data from industry to begin with. In lieu of national standards, it is time to look to regulatory alternatives to controlling the market, and—given their large negotiating power—the most promising seems to be AMCs setting higher policy standards.

#### V.B.1. AMC Commercialization and Use of Health Data and Biospecimens

A major value of private data and biobanks is in their ability to support good upstream research that can be translated into potentially lucrative downstream products with marketing authorization that clinicians will both use and prescribe. These kinds

---

247 PhRMA, *Code for Interactions with Healthcare Professionals* (2017).
248 Office of Inspector General, *Compliance Program Guidance for Pharmaceutical Manufacturers* (2003), [https://oig.hhs.gov/fraud/docs/complianceguidance/042803pharmacymfgnonfr.pdf](https://oig.hhs.gov/fraud/docs/complianceguidance/042803pharmacymfgnonfr.pdf) (accessed Apr. 2, 2018).
249 Pfizer Inc., *Corporate Integrity Agreement Annual Report* (2008).
250 Future of Privacy Forum, *supra* note 240 (‘The Best Practices provide a policy framework for the collection, retention, sharing, and use of Genetic Data generated by consumer genetic and personal genomic testing services.’).
251 Kayte Spector-Bagdady, *Improving Commercial Genetic Data Sharing Policy, in Consumer Genetic Technologies* (I. Glenn Cohen, Nita A. Farahany, Henry T. Greely, Carmel Shachar eds., Cambridge Univ. Press, forthcoming).
252 Maurer, *supra* note 234, at 188.
253 Spector-Bagdady, *Genetic data partnerships, supra* note 134.
254 Dinerstein, 2020 WL 5296920 at 1109 (‘Whatever a perpetual license for “Trained Models and Predictions” actually means, it appears to qualify as direct or indirect remuneration.’).
of comprehensive databanks require a diversity of both genomic and other health-related data, often found in EMRs. EMRs are generally in possession of hospitals and clinics providing health services.\(^{255}\) This gives AMCs negotiation power—not just for licenses for eventual machine-learning products such as in *Dinerstein v. Google*,\(^{256}\) but also to protect and improve the treatment of their patients and participants. Industry also needs AMCs to conduct and engage in research, publish articles, and treat the patients and write the prescriptions that make the private data valuable in the first place. Although academia can also benefit from using data and specimens held privately (as it might be more cost-effective for federally-funded researchers to purchase such data resources than generate them *de novo*\(^{257}\)), instead of waiting for industry to self-regulate its *production* of valuable health data and biospecimens, academia should self-regulate its own *consumption*.

In addition to the opportunity to potentially set better standards for the future, academia might also be motivated to set policies for engagement with private industry given the recent proliferation of related negative press coverage and lawsuits.\(^{258}\) Nontransparent academic/industry partnerships can bring attention to that which is legal, but not widely known (eg the Mayo/Google deal), or that which is unquestionably legal to begin with (eg the Google/Ascension deal). Both types of engagement raise questions of potential corruption, which can in turn ‘undermine[] the institution’s effectiveness by diverting it from its purpose or weaken[] its ability to achieve its purpose, including . . . the public’s trust in that institution or the institution’s inherent trustworthiness.’\(^{259}\)

But *partnerships* between industry and entities hoping to remain publicly oriented might raise a specter of corruption. As Jonathan Marks recently argued in his book, *The Perils of Partnership*, the focus of industry is profit and branding. This focus influences most, if not all, of industry’s actions—which may ‘lead to a bias toward the development of technological solutions to public health problems that may be readily commercialized.’\(^{260}\) For example, when soft drink companies partnered with public health agencies to jointly combat the ‘obesity epidemic’, the clinical focus was shifted from the effects of sugar to those of exercise; ie instead of focusing on limiting sugar intake, the proffered industry/government partnership solution was increasing physical activity to burn it off.\(^{261}\) Such partnerships not only allow industry to align government efforts on goals congruent with their bottom line, but also allow them to don a ‘health halo’ of respectability due to the assumption that government entities are acting in the public’s best interest—and those that they partner with do the same.\(^{262}\)

\(^{255}\) Price, *Contextual Bias*, supra note 101.
\(^{256}\) *Dinerstein*, 2020 WL 5296920 at 1109.
\(^{257}\) Eisenberg, *Noncompliance*, supra note 224, at 1086 (‘As long as it is cheaper for the owner to share the resource than it is for the user to recreate it, there are potential gains from exchange that stand to be dissipated through transaction costs or lost through bargaining breakdowns.’).
\(^{258}\) Eg Ross, supra note 127; Daisuke Wakabayashi, *Google and the University of Chicago Are Sued Over Data Sharing*, *NY Times*, June 26, 2019 https://www.nytimes.com/2019/06/26/technology/google-university-chicago-data-sharing-lawsuit.html (accessed July 18, 2020).
\(^{259}\) Lawrence Lessig, ‘*Institutional corruption* defined’, 41 J. LAW MED. ETHICS 553 (2013).
\(^{260}\) Marks, supra note 28, at 108.
\(^{261}\) Id. at 114.
\(^{262}\) Id. at 153.
We could also see this phenomenon painfully playing out in the initial glacial distribution of the COVID-19 vaccine across the country. States had been distributing their allocation of the vaccine from the Strategic National Stockpile to both public health entities as well as hospitals, AMCs, and other types of private entities.\textsuperscript{263} Although this may have made sense given the existing lack of infrastructure and funding of public health agencies,\textsuperscript{264} backlash regarding the choices of some of those entities—and in particular AMCs which adopted rather luxurious definitions of who counted as an ‘essential employee’—was swift.\textsuperscript{265} But this is a classic example of the problem with entrusting private entities with public goods—they are neither enabled (nor particularly inclined) to act equitably at a public community level. As Wendy Parmet recently argued in her \textit{Atlantic} essay on the privatization of public health:

Unquestionably, the private sector has a role to play in public health—just look at the private companies that produced the vaccines and the private hospitals that have cared for the ill. But to rely on it to protect the public’s health is pure folly. . . . To depend now on the private sector to increase vaccination rates would further underscore America’s tepid commitment to the basic principles of public health.\textsuperscript{266}

Thus, AMC/industry partnerships might not be the right framing to control data and biobank repositories. However, this should not prevent academia from setting standards for industry interactions. In fact, academia has done so successfully in the past. For example, in the wake of research establishing that even small gifts from pharmaceutical sales representatives influenced physician prescribing practices, many AMCs set policies limiting or prohibiting sales representatives from campus.\textsuperscript{267} These prohibitions, in turn, significantly altered prescribing practices (from predominantly drugs that had been heavily marketed to cheaper ones that were not) at the majority of implementing institutions.\textsuperscript{268} Setting standards at the AMC level is one distinct possibility for controlling these relationships.\textsuperscript{269}

As an example, at the University of Michigan Medical School, we have set up a multidisciplinary Human Data and Biospecimen Release Committee which reviews all industry and researcher requests to commercialize participant health data and

\begin{thebibliography}{99}
\bibitem{263} Deirdre Fernandes, \textit{Northeastern University starts vaccinating its front-line staff}, \textit{The Boston Globe} (Jan. 6, 2021) https://www.bostonglobe.com/2021/01/06/metro/northeastern-university-starts-vaccinating-its-front-line-staff/ (accessed Feb. 14, 2021).
\bibitem{264} Noah Higgins-Dunn, \textit{States will need billions to distribute the Covid vaccine as federal funding falls short}, \textit{CNBC} (Oct. 27, 2020) https://www.cnbc.com/2020/10/27/states-will-need-billions-to-distribute-the-covid-vaccine-as-federal-funding-falls-short.html (accessed Feb. 14, 2021).
\bibitem{265} Apoorva Mandavilli, \textit{At Elite Medical Centers, Even Workers Who Do not Qualify Are Vaccinated}, \textit{NYTimes} (Jan. 10, 2021) https://www.nytimes.com/2021/01/10/health/coronavirus-hospitals-vaccinations.html (accessed Feb. 14, 2021).
\bibitem{266} Wendy E. Parmet, \textit{Employers’ Vaccine Mandates Are Representative of America’s Failed Approach to Public Health}, \textit{Atlantic} (Feb. 4, 2021) https://www.theatlantic.com/ideas/archive/2021/02/privatization-public-health/617918/ (accessed Feb. 12, 2021).
\bibitem{267} Marks, supra note 28, at 30.
\bibitem{268} Ian Larkin et al., \textit{Association Between Academic Medical Center Pharmaceutical Detailing Policies and Physician Prescribing}, 317 JAMA 1785 (2017).
\bibitem{269} Kayte Spector-Bagdady et al., \textit{Sharing Health Data and Biospecimens with Industry: A Principle-Driven, Practical Approach}, 382 N. ENGL. J. MED. 2072 (2020).
\end{thebibliography}
biospecimens collected at Michigan Medicine. This Committee sets higher standards than those included in the human subjects research regulations, which we also voluntarily extend to all human subjects research at our institution. We require participant consent (with exceptions for rare diseases and some pediatric research) for all commercialization, and do not grandfather in data and biospecimens collected before the regulations required us to disclose this information. In addition, given the increasing risk of reidentification discussed above, we require commercialization consent of even data and biospecimens that will be provided to industry in deidentified form.

Another example of a potential AMC approach is controlling the data usage agreements necessary to transition industry data to academic researchers at most AMCs. AMCs could require additional clarity regarding industry data provenance specifically to ensure that the type of informed consent provided by participants meets the AMC’s own expectations; avoiding the opportunity for AMC researchers to otherwise white-wash data that they could not have legally acquired themselves. Indeed, there has been minimal oversight or even acknowledgement of the consent pedigree of privately procured specimens and data in the published literature. And, although there has been some response to the moral turpitude of analyzing specimens and data taken from participants forcefully—such as via the Guatemala STD experiments, Nazi medical experiments, or Chinese prisoners—there has been virtually no emphasis in nonegregious circumstances. For example, in a recent review of academic publications with data procured from private databanks, the type of original contributor consent was either not disclosed or was unclear almost half the time. AMCs could begin to improve this system at the data use agreement level.

Another possibility is that AMCs could require IRB authorization of the primary research which generated the data in the first place. While, as discussed above, IRB authorization might not be an actual legal requirement of all industry research, by refusing to publish secondary research with data acquired otherwise, AMCs could begin to shift behavior. Such an approach has been employed by journal editors in the past. For example, when 23andMe submitted one of its first genome-wide association studies to PLoS Genetics in 2010, editors noticed three major deficits: the protocol had not been prospectively reviewed by an IRB, there was concern over the type of consent obtained, and access to the underlying data was limited. PLoS Genetics editors published a concurrent comment explaining why they decided to publish the piece, despite these perceived shortcomings, given the work’s (i) novel contribution to the literature and (ii)

270 Id.
271 Id.
272 Michelle M. Mello et al., Waiting for Data: Barriers to Executing Data Use Agreements, 367 SCIENCE 150 (2020).
273 Kayte Spector-Bagdady & Paul A. Lombardo, From in vivo to in vitro: How the Guatemala STD Experiments Transformed Bodies into Biospecimens, 96 MILBANK Q. 244 (2018).
274 Stephen G. Post, The Echo of Nuremberg: Nazi Data and Ethics, 17 J. MED. ETHICS 42 (1991).
275 Retraction Watch, Journals retract more than a dozen studies from China that may have used executed prisoners’ organs, Aug. 14, 2019, https://retractionwatch.com/2019/08/14/journals-retract-more-than-a-dozen-studies-from-china-that-may-have-used-executed-prisoners-organs/ (accessed July 18, 2020).
276 Spector-Bagdady, Genetic data partnerships, supra note 134.
277 Nicholas Eriksson et al., WebBased, Participant Driven Studies Yield Novel Genetic Associations for Common Traits, 6 PLoS GENET. E1000993 (2010).
their independent assessment that the participants involved were neither coerced nor deceived. Ultimately, editors argued that publication ‘accompanied by an editorial providing transparent documentation of the process of consideration’ was appropriate, in addition to their ‘call for community input to spur efforts to standardize the IRB consent process’ for that type of research. The day the editorial was published, 23andMe sent out a press release emphasizing that IRB review was not legally required for its research—but that it would obtain review going forward ‘to ensure that our work is in line with scientific research best practices.’ In this case, the ability to publish knowledge gained from analysis of its databank made the databank a more valuable business asset itself—one for which 23andMe was willing to give up something of value (ie not having IRB oversight) in return.

V.B.2. ‘Lifting All Boats’ Via Informed Consent

In addition to better controlling the flow of data and specimens between AMCs and industry, AMCs can begin to improve informed consent practices at their own institution. As argued above, one of the reasons that the information contained in a consent form is so highly regulated is that it is easier than empirically validated methods of gaining informed consent (ie a conversation) to regulate in the first place. The assumption is that if information about risks or burdens were disclosed in the form, they have been understood and accepted. We know this is not true. But, given the new types of risks to individuals and benefits to communities of secondary research, should we bother trying to continue to improve the individual informed consent process given the enormous time and burden that would entail? If the individual risks of secondary research are less, and we know that informed consent is generally ineffective anyway, is there a better solution?

Although a signature on the informed consent form will certainly remain legally required, AMCs can also attempt to better conscribe what they are asking of patients and participants contributing to secondary research banks in the first place. By ensuring a baseline standard taking into account the risks and benefits of secondary research, with buy in from representative community members, this ‘rising tide can lift all boats’ and potentially better protect many more participants—even if some might not read or understand the forms.

One way to do this is by establishing some type of data review committee which will agree to prospective standards for what can be asked of participants in the first place, rather than a retrospective check of informed consent (like the University of Michigan

278 Greg Gibson & Gregory P. Copenhaver, Consent and Internet-Enabled Human Genomics, 6 PLOS GENET. E1000965 (2010).

279 Id.

280 23andMe, 23andMe improves research consent, Jun. 24, 2010, https://blog.23andme.com/23andme-research/23andme-improves-research-consent-process/ (accessed Mar. 31, 2019).

281 Kayte Spector-Bagdady, ‘The Google of Healthcare’: Enabling the Privatization of Genetic Bio/Databanking, 26 ANN. EPIDEMIOLOG. 515, 517 (2016).

282 As Beecher argued as far back as 1966: ‘All so-called codes are based on the bland assumption that meaningful or informed consent is readily available for the asking . . . this is very often not the case. Consent in any fully informed sense may not be obtainable. Nevertheless, except, possibly, in the most trivial situations, it remains a goal toward which one must strive for sociologic, ethical and clear-cut legal reasons. There is no choice in the matter.’ Beecher, supra note 33, at 1355.
model). Not only would such a committee potentially resolve the imbalance of putting too much emphasis on individual consent and autonomy for low risk research, but it would also avoid the secondary issue that even when researchers attempt to get individual level consent, participants rarely comprehend what they are agreeing to.

At the committee level, patient representatives could engage in a more helpful conversation about risks, benefits, and alternatives to research. In particular, given that we know that Black and Latino participants are more likely to be hesitant to share data and specimens, we must ensure that these standards are not set at the ‘average patient’ standard—because that ‘average’ for many AMCs is likely to be of a prominently white cohort and therefore represent a white viewpoint. Informed consent standards should be responsive to a diversity of racial and ethnic viewpoints so as not to further discourage the diversity of racial and ethnic representation within research datasets.

Of course, dissenting individual participants would still be potentially subjugated via representation at the cohort level. But, given the justice and public beneficence opportunity for communities from enabling such research, and the low risk to individuals, this seems like a better compromise than the one we are making now (i.e. that many participants do not understand the forms and yet we move forward). We should instead be moving toward setting and enforcing higher standards across the board such that the individual informed consent conversation can focus on other preference-sensitive choices (e.g., amount of time participants are willing to commit to research in return for what value).

Therefore, although a more classic academic/industry partnership model might not be appropriate, there are several concrete steps AMCs can take to better control and set standards for their own data and specimen commercialization and academic researcher access to industry data and specimens, as well as resolve some of the lapses in protection of patients and participants at the federal level.

VI. CONCLUSION
The future of scientific advances via secondary research with biospecimens and health data is bright. However, the current strict governance of federally funded researchers (and, by association, most AMC researchers) and a decided lack of governance of privately collected data and specimens has created a stark imbalance. This inequity is increasingly leading to AMCs commercializing their own health data and specimens, in addition to securing additional data access from private holdings. But, though AMCs may have little control over new federal legislation or revising extant regulations, they do have control over supply and demand as well as the behavior of their researcher employees. Simply succumbing to inevitable privatization of data and biospecimen banking is not an acceptable solution for such an important public good. Instead, AMCs should move toward better controlling the proverbial revolving door between themselves and industry to continue to advance life-saving research while also protecting the interests of those whose data are supporting such advances.

283 Jagsi et al., supra note 9.