
ARTICLE

GOVERNING HEALTH INFORMATION

CRAIG KONNOTH[†]

In the last few years, numerous Americans' health information has been collected and used for follow-on, secondary research. This research studies correlations between medical conditions, genetic or behavioral profiles, and treatments, to customize medical care to specific individuals. Recent federal legislation and regulations make it easier to collect and use the data of the low-income, unwell, and elderly for this purpose. This would impose disproportionate security and autonomy burdens on these individuals. Those who are well-off and pay out of pocket could effectively exempt their data from the publicly available information pot. This presents a problem which modern research ethics is not well equipped to address. Where it considers equity at all, it emphasizes underinclusion and the disproportionate distribution of research benefits, rather than overinclusion and disproportionate distribution of burdens.

I rely on basic intuitions of reciprocity and fair play as well as broader accounts of social and political equity to show that equity in burden distribution is a key aspect of the ethics of secondary research. To satisfy its demands, we can use three sets of regulatory and policy levers. First, information collection for public research should expand beyond groups having the lowest welfare. Next, data analyses and queries should draw on data pools more equitably. Finally, we must create an entity to coordinate these solutions using existing statutory authority if possible. Considering health information collection at a systematic level—rather than that of individual

[†] Sharswood Fellow and Lecturer in Law, University of Pennsylvania Law School; Rudin Fellow, NYU Langone Medical Center. J.D., Yale Law School; M.Phil., University of Cambridge. My thanks to Guy Charles, Abbe Gluck, Anita Allen, Frank Pasquale, Nicolas Terry, Regina Austin, Stephanos Bibas, Glenn Cohen, Leslie Francis, Jonah Gelbach, Daniel Hemel, Allison Hoffman, Serena Mayeri, Bill Sage, Dara Purvis, Christopher Robertson, Lindsay Wiley, Tess Wilkinson-Ryan, and participants of the Duke University School of Law Emerging Scholars Workshop, William & Mary School of Law Faculty Workshop, and St. Louis University School of Law Health Scholars Workshop. Thanks are owed to Taylor Goodspeed and Rick Mula for research assistance and Alvin Dong for library assistance.

clinical encounters—gives us insight into the broader role that health information plays in forming personhood, citizenship, and community.

INTRODUCTION	1318
I. INEQUITY IN HEALTH INFORMATION COLLECTION	1323
A. <i>Low-Income and Elderly Individuals</i>	1325
B. <i>The Privately Insured</i>	1329
C. <i>High-Income Individuals</i>	1331
II. THE BURDENS OF HEALTH INFORMATION COLLECTION	1333
A. <i>Privacy and Security Harms</i>	1334
B. <i>Personhood Harms</i>	1338
C. <i>Expressive Harms</i>	1343
D. <i>Medical Harms</i>	1346
III. ARGUING FOR EQUITY	1348
A. <i>Health Information Norms</i>	1351
B. <i>General Social Norms</i>	1354
C. <i>Correcting Inequity</i>	1357
D. <i>A Utilitarian Alternative</i>	1360
IV. USING LAW TO ACHIEVE EQUITY	1362
A. <i>Collection</i>	1362
1. <i>EHR Data</i>	1362
2. <i>Claims Data</i>	1364
3. <i>The HIPAA Exclusion</i>	1366
4. <i>Data Collection from the Healthy</i>	1366
5. <i>Other Data Collection Policies</i>	1368
B. <i>Research</i>	1369
C. <i>Coordination</i>	1371
CONCLUSION	1375

INTRODUCTION

Medical research is undergoing its third paradigm shift of the last three centuries.¹ Until the nineteenth century, research mostly occurred informally in the course of treating a patient and observing outcomes—a physician

¹ The approach I describe “highlights a historic shift in medical research and care . . . This type of study marks a shift not just in the way doctors treat patients but also in how they conduct long-term medical research.” Steve Sternberg, *The Precision Medicine Revolution*, U.S. NEWS & WORLD REP. (June 6, 2016, 9:00 AM), <http://www.usnews.com/news/articles/2016-06-06/precision-medicine-is-changing-the-nature-of-long-term-medical-research> [<https://perma.cc/9TK7-VDDY>].

learned by doing.² Subsequently, the clinical trial became the staple of medical research in the twentieth century.³ The clinical trial remains important in the twenty-first century, but breakthroughs are increasingly coming from “informational” or “secondary” research. By this, I mean research that aggregates information about patients, including physical conditions, genetic information, treatments, responses, and outcomes.⁴ This type of research provides a real-world snapshot at a population-wide level in ways that are not possible with traditional clinical trials. Data from clinical contexts are fed back into databases in a “continuous feedback loop,” the analysis of which iteratively helps to improve clinical and health delivery outcomes.⁵ The goal is to create a national information network based on data collected from providers, payers, or patients by private or public entities.

The collection of such information raises serious ethical concerns because it imposes special burdens on specific patients whose records form the data pool for queries and analyses. Even with the best protections, “[n]o security measures . . . can ever completely safeguard against . . . release of . . . or inappropriate use of information.”⁶ Patients therefore face the risk of

² See Emily A. Largent et al., *Can Research and Care Be Ethically Integrated?*, 41 HASTINGS CTR. REP., July–Aug. 2011, at 37, 37 (describing the historical practice of carrying out research in the course of providing medical care).

³ See Barbara J. Evans, *Much Ado About Data Ownership*, 25 HARV. J.L. & TECH. 69, 76 (2011) (recognizing that “randomized, controlled clinical trials . . . were the major workhorse of late twentieth-century biomedical discovery”).

⁴ While my usage is standard, other scholars prefer different terms. See, e.g., SHARYL J. NASS ET AL., INST. OF MED. OF THE NAT’L ACADS., BEYOND THE HIPAA PRIVACY RULE: ENHANCING PRIVACY, IMPROVING HEALTH THROUGH RESEARCH 19 & n.11 (2009) [hereinafter BEYOND THE HIPAA PRIVACY RULE] (“The National Committee on Vital and Health Statistics has noted that the term ‘secondary uses’ of health data is ill defined and therefore urged abandoning it . . .”). This Article is solely about secondary research based on data that is identifiable so that records about a single patient that were collected from multiple sources and at different points in time are linkable. See Sharona Hoffman & Andy Podgurski, *Balancing Privacy, Autonomy, and Scientific Needs in Electronic Health Records Research*, 65 SMU L. REV. 85, 130 (2012) (describing a study that concluded that removing identifier elements from medical data in compliance with HIPAA’s “de-identif[cation]” rules “reduced data by 31% and precluded access to information that is vitally important for research purposes”). The Article does not concern research on biospecimens or clinical trials.

⁵ JOE ALPER & CLAUDIA GROSSMAN, INST. OF MED. OF THE NAT’L ACADS., INTEGRATING RESEARCH AND PRACTICE: HEALTH SYSTEM LEADERS WORKING TOWARD HIGH-VALUE CARE 13 (2015) [hereinafter INTEGRATING RESEARCH AND PRACTICE]. Here, I rely on the traditional Common Rule definition of research as “a systematic investigation, including research development, testing and evaluation, designed to develop or contribute to generalizable knowledge.” 45 C.F.R. § 46.102(d) (2016). Thus, I envisage research as contributing to the social good, broadly defined. The definition of research is a complex subject involving academic dispute and discussion that is beyond the scope of this Article.

⁶ KATHY HUDSON ET AL., THE PRECISION MED. INITIATIVE WORKING GRP., PRECISION MEDICINE INITIATIVE COHORT PROGRAM—BUILDING A RESEARCH FOUNDATION FOR 21ST CENTURY MEDICINE 84 (2015), <https://acd.od.nih.gov/reports/DRAFT-PMI-WG-Report-9-11->

employment or insurance discrimination, reputational loss, or identity theft.⁷ Individuals also have personal interests in their health information, whether identified or de-identified, much as they do with property.⁸ They suffer dignitary harm when this information is used without their consent. Further, surveillance of, and information collection from certain groups, can send certain social messages about them that affect their status. Finally, these harms, whether real or perceived, may reduce trust in the medical system generally, exacerbating health problems among those groups.

The key problem is this: new regulations and laws increasingly facilitate the collection and public use of data of those on public benefit programs, namely, the poor and the elderly. By contrast, others who do not rely on public benefit programs can keep their information out of the communal pot while remaining well-positioned to enjoy the health benefits that follow as the learning health system gets off the ground.

This Article argues that laws should distribute information burdens across society in a just manner.⁹ This entails taking into account the social welfare of the individual patient where possible when imposing information burdens. In concrete terms, this requires altering the points at which we collect information that is made publicly available, focusing less on public benefit programs like Medicare and Medicaid, and looking to other sources of data. We must also alter research methods by broadening and—where possible—shifting, the public data pool that is queried for research. Many of the solutions that I offer can be realized through simple regulatory changes. I also offer statutory solutions, although recognizing that those may be far more difficult to achieve.

Nonetheless, the law often, and with good reason, imposes different material, dignitary, and autonomy burdens on different groups. The Constitution and existing statutes provide little basis to suggest changes to how informational burdens are distributed. However, bioethical precepts, which have historically shaped research laws and regulations,¹⁰ require a more equitable burden allocation. Accordingly, this Article develops a framework grounded in bioethics to support health information equity.

2015-508.pdf [https://perma.cc/7WHK-NWPU] [hereinafter THE PRECISION MEDICINE INITIATIVE]. The harms can range from reidentification to the misuse of one's private information.

⁷ See M. Ryan Calo, *The Boundaries of Privacy Harm*, 86 IND. L.J. 1131, 1148 (2011) (noting the wide concern of privacy invasion leading to "adverse, real-world consequence[s]"); see also Craig Konnoth, *An Expressive Theory of Privacy Intrusions*, 102 IOWA L. REV. (forthcoming 2017) (manuscript at 9) (on file with author) ("Privacy intrusions impose status harms on individuals at the time of the intrusion.").

⁸ See *infra* notes 109–31 and accompanying text.

⁹ I note here that my critique is only of the laws as written. In practice, the effects of the law could be counteracted by other forces that are not yet predictable.

¹⁰ See *infra* note 153 and accompanying text.

The ethics of health information research have largely been dominated by a single value—maintaining individual autonomy in clinical encounters—and offer mechanisms such as individual consent and privacy protections to preserve this autonomy.¹¹ This approach offers little purchase for equity concerns. To be sure, bioethicists reminded us to distribute the burdens of research equitably a generation ago.¹² But those admonishments were offered with little analysis. And in the years since, research ethics have undergone a “dramatic change.”¹³ Today, the field focuses almost exclusively on distributing the *benefits* of research and on questions of *underinclusion* of minorities—rather than *overinclusion*.¹⁴

11 See Nicolas P. Terry, *Big Data Proxies and Health Privacy Exceptionalism*, 24 HEALTH MATRIX J.L.-MED. 65, 99-100 (2014) (describing the traditional autonomy-centered model of health information in which physicians are bound by a “duty of confidentiality in curating the[ir] patient’s health data”); Nicolas P. Terry, *Protecting Patient Privacy in the Age of Big Data*, 81 UMKC L. REV. 385, 401 (2012) [hereinafter Terry, *Protecting Patient Privacy*] (defining health privacy exceptionalism as the idea that “health information deserves a higher level of privacy protection than most other types of data”); see also Hoffman & Podgurski, *supra* note 4, at 91-94 (noting that many parties are interested in acquiring the sensitive and personal information from medical records); Sharona Hoffman & Andy Podgurski, *In Sickness, Health, and Cyberspace: Protecting the Security of Electronic Private Health Information*, 48 B.C. L. REV. 331, 339 (2007) (listing the requirements of the HIPAA Security Rule for covered entities, which include “protect[ing] [health] data against reasonably anticipated threats to its security”); Peter D. Jacobson, *Medical Records and HIPAA: Is It Too Late to Protect Privacy?*, 86 MINN. L. REV. 1497, 1497-99 (2002) (recognizing that medical records are “highly protected” under HIPAA and that privacy is a fundamental value to the health care enterprise); Nicolas P. Terry, *Electronic Health Records: International, Structural and Legal Perspectives*, 12 J.L. & MED. 26, 29 (2004) (explaining that “patient interests in confidentiality, privacy and anonymity” generate tensions “between the key stakeholders and their needs” in the health information technology domain); Nicolas P. Terry & Leslie P. Francis, *Ensuring the Privacy and Confidentiality of Electronic Health Records*, 2007 U. ILL. L. REV. 681, 689-90 (describing the Australian *HealthConnect* system, which “allows the patient (in consultation with the physician) to control what data are included and who may view it”).

12 See Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research, Report of the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 44 Fed. Reg. 23,192, 23,192, 23,194 (Apr. 18, 1979) [hereinafter Belmont Report] (naming “justice” as a basic ethical principle particularly relevant to the ethics of research of human subjects and explaining that “an injustice occurs . . . when some burden is imposed unduly”).

13 Patricia A. King, *Justice Beyond Belmont*, in BELMONT REVISITED: ETHICAL PRINCIPLES FOR RESEARCH WITH HUMAN SUBJECTS 136, 136 (James F. Childress et al. eds., 2005).

14 Indeed, scholars suggest that the key problem is that low-income populations are underincluded in databases. See Jonas Lerman, *Big Data and Its Exclusions*, 66 STAN. L. REV. ONLINE 55, 56 (2013) (“[T]he reality is that billions of people remain on its margins because they do not routinely engage in activities that big data and advanced analytics are designed to capture.”); Sarah E. Malanga et al., *Big Data Neglects Populations Most in Need of Medical and Public Health Research and Interventions* (Ariz. Legal Studies, Discussion Paper No. 16-26) in BIG DATA, HEALTH LAW, AND BIOETHICS (H.F. Lynch et al. eds., forthcoming 2017) (manuscript at 7), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2828954## (arguing that underserved communities, such as racial and ethnic minorities, have been excluded from many data sets and warning that such exclusion

Reclaiming equity requires us to assess the obligations the players within the system can place on each other.¹⁵ I first consider these obligations as if the health information system were a closed world. There, simple, intuitive norms of reciprocity require a greater degree of equity. The overwhelming majority of individuals will benefit from health information over their lifetimes. To conscript the information of and impose material, dignitary, and status harms on only one group undermines fundamental notions of fair play.

But the health information system is also embedded in a larger set of social and political institutions against which these equity values should be tested. Different approaches to these broader institutions apply different values and obligations throughout society. These society-wide prescriptions may influence the obligations within particular systems, including that of health information. I thus take more complex, but nonetheless widely shared intuitions as given: that the poor and elderly deserve assistance, that in existing society, the poor and perhaps, the elderly, do not receive the level of assistance they deserve, and that addressing health information burdens will alleviate some of these other social harms. I recognize though, that not everyone shares these theories of obligation. But this is not the place to defend them at length. I simply show how altering each of my premises affects my claims about equity to different degrees.

Part I lays out how new laws and policies are facilitating the disproportionate collection and public use of data. Part II details the kinds of burdens such practices can impose. Part III provides an ethical framework to assess these inequities. Part IV then shows what regulatory and statutory levers can be used to render secondary research more equitable.

Finally, I emphasize that this Article does not address the question of *whether* we should collect information—a question which already dominates the information ethics literature.¹⁶ Rather, it accepts as a premise that it is

limits the extent to which health research will benefit these groups). In doing so, they do not mention many of the developments that I discuss in this Article.

¹⁵ I assume here that identifying “obligations” justifies government coercion of individuals into complying with those obligations.

¹⁶ A prominent line of research ethics argues that there is an obligation to participate in research. See, e.g., Arthur L. Caplan, *Is There a Duty to Serve as a Subject in Biomedical Research?*, 6 IRB: ETHICS & HUM. RES., Sept./Oct. 1984, at 1, 3 (suggesting that those who benefit from research have an obligation to participate in research); John Harris, *Scientific Research Is a Moral Duty*, 31 J. MED. ETHICS 242, 242 (2005) (citing two principles, “do no harm” and “fairness,” as the sources of a “powerful obligation to pursue, support, and participate in scientific research” (capitalization omitted)). Even assuming such a duty, questions would still remain about the extent of that duty, and privacy ethicists argue for—and the law recognizes—various kinds of limitations. See Craig Konnoth, *Classification Standards for Health Information: Ethical and Practical Approaches*, 72 WASH. & LEE L. REV. ONLINE 395, 404 (2016) (proposing “a more calibrated breadth of consent” in which individuals consent to their information being used “only for research that departs from the original context of collection [to] a certain [extent],” as distinct from the current “all-or-nothing

valid to collect health information to serve the numerous goals of the health information system. Notwithstanding that premise, we must adopt a framework within which to reorganize privacy risk in ways that are ethical and just. Where bioethics has sought only to incorporate autonomy concerns in health data collection, this framework provides a guide for moving beyond autonomy to equity concerns.

I. INEQUITY IN HEALTH INFORMATION COLLECTION

The benefits of secondary research promise to be numerous. Agglomerating health records has enabled researchers to identify genetic mutations that indicate a high risk of breast cancer or Alzheimer's,¹⁷ make changes to drug choice and administration,¹⁸ and develop quality and cost control measures.¹⁹ Secondary research promises to battle discrimination and stigma of certain kinds by revealing health care disparities across populations and the prevalence of certain conditions.²⁰ It facilitates recruitment for clinical trials²¹ and enables other

approach . . . [to] provide broad consent for all further research"). *But see* W. Lipworth et al., *Consent in Crisis: The Need to Reconceptualize Consent to Tissue Banking Research*, 36 INTERNAL MED. J. 124, 125 (2006) (noting that more stringent consent requirements may prove to be a burden on researchers by "creat[ing] an untenable workload and reduc[ing] sample sizes").

¹⁷ See *Ass'n for Molecular Pathology v. Myriad Genetics, Inc.*, 133 S. Ct. 2107, 2117 (2013) (explaining that Myriad Genetics "found the location of a gene associated with increased risk of breast cancer and identified mutations of that gene that increase the risk"); *Alzheimer's Genes: Are You at Risk?*, MAYO CLINIC, <http://www.mayoclinic.org/diseases-conditions/alzheimers-disease/in-depth/alzheimers-genes/art-20046552> [<https://perma.cc/XU6C-5ZAK>] ("Researchers have identified a number of genes associated with Alzheimer's disease."). For a longer list of genetic diagnoses, see Fred H. Cate, *Protecting Privacy in Health Research: The Limits of Individual Choice*, 98 CALIF. L. REV. 1765, 1780 (2010), which discusses mutations that contraindicate the use of the popular blood thinner, warfarin, and *Preimplantation Genetic Diagnosis*, GENESIS GENETICS, <http://genesigenetics.org/pgd/> [<https://perma.cc/V3ZB-CBXA>].

¹⁸ See Allen D. Roses, *Pharmacogenetics and Future Drug Development and Delivery*, 355 LANCET 1358, 1358 (2000) (discussing the role of genetic information in medication decisions).

¹⁹ See, e.g., ALEX PENTLAND, TODD G. REID & TRACY HEIBECK, *BIG DATA AND HEALTH: REVOLUTIONIZING MEDICINE AND PUBLIC HEALTH* 31 (2013) (stating that readmissions correlated with mental depression in Washington D.C. hospitals); *id.* at 30 (analyzing unnecessary prescriptions of expensive brand name medication); see also Lawrence O. Gostin, *Health Information: Reconciling Personal Privacy with the Public Good of Human Health*, 9 HEALTH CARE ANALYSIS 321, 322 (2001) (advocating for the collection of health data for, among other things, the facilitation of cost-effective assessment).

²⁰ See Gregory E. Simon et al., *Large Medical Databases, Population-Based Research, and Patient Confidentiality*, 157 AM. J. PSYCHIATRY 1731, 1732 (2000) ("Data documenting adverse effects of older antidepressant drugs and long-acting sedative-hypnotic drugs have informed efforts to improve the quality of psychotropic drug prescribing for older adults . . .").

²¹ See Tracy Stuardi, Helen Cox & David J. Torgerson, *Database Recruitment: A Solution to Poor Recruitment in Randomized Trials?*, 28 FAM. PRAC. 329, 330 (2010) (discussing the advantages of database recruitment for randomized controlled trials).

means of research when trials are not possible.²² Over the summer of 2016, the then–National Coordinator for Health Information Technology observed at an internal medicine research review session that “5 out of 6 [studies] used [electronic health record] data.”²³ Another study finds that three-fifths of clinical trials use some form of digital health information.²⁴

Only some health data makes it into secondary research because not all individuals go to the doctor—either because they are healthy or because they cannot afford health care (but are not poor enough to receive Medicaid).²⁵ A doctor may also fail to enter information into a health record or transmit gathered data.

The data that is captured and used are of two main varieties. First, there is claims data that providers transfer to payers to receive reimbursement. This includes diagnostic and procedure codes.²⁶ Next, there is electronic health or medical record (EHR) data. While claims data only contains information about a provider’s conclusions (i.e., her diagnoses) and treatment procedures,²⁷ the EHR contains the information on which the conclusion was based—such as symptoms, behavior, and even genetic information.²⁸ Claims data is mainly used to assess health care delivery, cost and utilization, and sometimes, quality

22 See generally Walter F. Stewart et al., *Bridging the Inferential Gap: The Electronic Health Record and Clinical Evidence*, 26 HEALTH AFF. w181 (2007) (explaining various shortcomings of randomized control trials, including that they are too selective and ignore comorbidities, and noting that secondary research helps to bridge the gap).

23 Darius Tahir, *Inspector General on DOD/Cerner Implementation*, POLITICO (June 2, 2016, 10:00 AM), <http://www.politico.com/tipsheets/morning-ehealth/2016/06/inspector-general-on-dod-cerner-implementation-unc-transparency-site-tiff-desalvo-on-pulse-check-214610> [https://perma.cc/YU8M-XNTY] (alteration in original) (internal quotation marks omitted).

24 See Validic, *Validic Reveals Findings from 2016 Global Pharma and Biotech Survey on Digital Health*, PR NEWSWIRE (Sept. 19, 2016, 8:16 PM), <http://www.prnewswire.com/news-releases/validic-reveals-findings-from-2016-global-pharma-and-biotech-survey-on-digital-health-300330012.html> [https://perma.cc/9A5Q-XK54] (“According to survey results, over 60 percent of respondents stated that they have used digital health technologies in clinical trials . . .”).

25 See JOHN WILSON & ADAM BOCK, OPTUM THE BENEFIT OF USING BOTH CLAIMS DATA AND ELECTRONIC MEDICAL RECORD DATA IN HEALTH CARE ANALYSIS 4 (2014), <https://www.optum.com/content/dam/optum/resources/whitePapers/Benefits-of-using-both-claims-and-EMR-data-in-HC-analysis-WhitePaper-ACS.pdf> [https://perma.cc/9XHJ-MLKP] (explaining that “uninsured patients have no source of claim data”).

26 See *HCPCS-General Information*, CMS.GOV, <https://www.cms.gov/medicare/coding/medhcpcsgeninfo/index.html> [https://perma.cc/63AV-A4PR] (providing background information about the Healthcare Common Procedure Coding System (HCPCS) that health insurance providers use to process claims).

27 *Id.*

28 *Electronic Health Records*, CMS.GOV, <https://www.cms.gov/Medicare/E-health/EHealthRecords/index.html> [https://perma.cc/8LQ4-MMMB] (last updated Mar. 26, 2012). As one doctor notes, “Whew! We think the ICD-10 [diagnostic] coding hierarchy is complex—it’s nothing compared to [the EHR coding mechanism] SNOMED-CT.” Eden Ware, *SNOMED: What It Is and Why It Was Added to Stage 2 Meaningful Use*, HEALTH LANGUAGE (Jan. 25, 2013), <http://blog.healthlanguage.com/SNOMED-What-it-is-and-Why-it-was-Added-to-Stage-2-Meaningful-Use> [https://perma.cc/94E5-NHLF].

of care.²⁹ It is less frequently used for clinical research. EHR data, on the other hand, is far more granular and time-limited, and can therefore be used for clinical research.

Both kinds of data can come from many sources. For example, patients can voluntarily provide their information to research networks. Alternatively, payers and providers, both public and private, can collect data and analyze it themselves or pass it on to other entities for analysis. Laws and policies have made data from public programs progressively more attractive and available for public research, while private data is increasingly harder to aggregate.

Although public programs historically made *claims* data publicly available for research, they did not collect *EHR* data. This meant that those seeking EHR data needed to rely on difficult-to-access private sources. However, public programs are ramping up EHR data collection. These changes are coinciding with increased enrollment in public programs and increased access to data—sometimes including names and social security numbers. On the other hand, the data of other individuals is being left outside the publicly available information pot.

A. Low-Income and Elderly Individuals

Data collection policies affecting low-income and elderly individuals are increasingly being determined by the convergence of two processes.³⁰ First, more and more data is being collected from individuals in public benefits programs who tend to be poorer, older, and in worse health.³¹ For example, within the last two years, the granularity of diagnostic codes—and therefore the details about beneficiaries' illnesses—has increased by more than a factor of seven.³² Further, the Center for Medicare and Medicaid Services (CMS)

²⁹ WILSON & BOCK, *supra* note 25, at 3.

³⁰ Medicaid includes only the poorest individuals. Those below 100% of the poverty threshold have the lowest rates of insurance coverage while those at or above 400% have the highest rates. Jessica C. Smith & Carla Medalia, U.S. CENSUS BUREAU, HEALTH INSURANCE COVERAGE IN THE UNITED STATES: 2014, at 13 (2015).

³¹ See Gretchen Jacobson et al., *Income and Assets of Medicare Beneficiaries, 2014–2030*, HENRY J. KAISER FAM. FOUND. (Sept. 10, 2015), <http://kff.org/medicare/issue-brief/income-and-assets-of-medicare-beneficiaries-2014-2030> [<https://perma.cc/L9PX-C4C7>] (compiling financial data for Medicare beneficiaries and finding that “most are of modest means”). Moreover, the health of the elderly tends to be worse. Furthermore, Medicaid is, of course, a means-tested program and Medicaid recipients are in “poorer health” than the privately insured. Julia Paradise & Rachel Garfield, *What Is Medicaid's Impact on Access to Care, Health Outcomes, and Quality of Care? Setting the Record Straight on the Evidence*, HENRY J. KAISER FAM. FOUND. (Aug. 2, 2013), <http://kff.org/report-section/what-is-medicaids-impact-on-access-to-care-health-outcomes-and-quality-of-care-setting-the-record-straight-on-the-evidence-issue-brief/> [<https://perma.cc/GX2P-KC96>].

³² The ICD-9 coding system, used until 2015, had under 20,000 unique codes. However, reporting parameters are becoming more granular. ICD-10, adopted in 2015, allows the reporting of close to 150,000 distinct codes. See *International Classification of Diseases, (ICD-10-CM/PCS) Transition-*

now mandates the use of electronic health records.³³ EHRs include data about clinical and physiological symptoms, as well as family, medical, environmental, and social details, including information about diet, employment, exercise, and alcohol habits.³⁴ Providers must also transmit certain aggregate measures to centralized registries based on this data,³⁵ with more granularity over time.³⁶ Additionally, the Affordable Care Act ensures that all of this data is collected from more individuals: in 2015, eleven million additional individuals became eligible for Medicaid.³⁷

Next, new policies offer “unprecedented” data access and concomitant risk.³⁸ Although Medicare and Medicaid claims data has long been available for public

Background, CDC, http://www.cdc.gov/nchs/icd/icd10cm_pcs_background.htm [<https://perma.cc/86SA-FEBM>] (last updated Oct. 1, 2015). ICD-10 also provides an opportunity to capture much greater specificity about the nature of a patient’s condition, which makes it more useful for research. See JOSEPH C. NICHOLS, HEALTH DATA CONSULTING, DATA VALUE: BREAKING OLD HABITS 4 (2016), <http://pages.himss.org/WZonIVLkGot3WRgookiWQoo> [<https://perma.cc/7X3A-PH87>] (“ICD-10 provides the ability to collect even greater detail about the location and type of breast cancer.”).

³³ See *EHR Incentive Payment Timeline*, HEALTHIT.GOV, <https://www.healthit.gov/providers-professionals/ehr-incentive-payment-timeline> [<https://perma.cc/S47K-GMT5>] (last updated Mar. 4, 2014) (“Financial penalties are scheduled to take effect in 2015 for Medicare and Medicaid providers who did not transition to EHRs.”). Despite rumors, the fundamentals of the program remain despite recent regulatory changes. *Electronic Health Records (EHR) Incentive Programs*, CMS.GOV, <https://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/index.html?redirect=/ehrincentiveprograms> [<https://perma.cc/H23Z-YMSU>] (last updated Feb. 8, 2017). As this Article goes to press, the public fate of these programs under the Trump Administration remains in limbo.

³⁴ A list of many of these variables in standardized format can be sought in the SNOMED library at *SNOMED International SNOMED CT Browser*, SNOMED CT, <http://browser.ihtsdotools.org/?perspective=full&conceptId1=404684003&edition=en-edition&release=v20170131&server=http://browser.ihtsdotools.org/api/snomed&langRefset=90000000000509007> [<https://perma.cc/ED72-PSCZ>].

³⁵ See Medicare and Medicaid Programs; Electronic Health Record Incentive Program—Stage 3 and Modifications to Meaningful Use in 2015 Through 2017, 80 Fed. Reg. 62,762, 62,818 (Oct. 16, 2015) (to be codified at 42 C.F.R. pts. 412, 495) (announcing a modified reporting requirement concerning the reporting of electronic public health data to clinical data registries).

³⁶ Clinical Quality Measures (CQMs) usually require reporting of (patients with a certain condition) / (all patients in certain demographic or clinical group). *eCQM Library*, CMS.GOV (last updated Mar. 21, 2017), https://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/eCQM_Library.html [<https://perma.cc/H6W5-X4QN>].

³⁷ Robin Rudowitz et al., *What Coverage and Financing Is at Risk Under a Repeal of the ACA Medicaid Expansion*, HENRY J. KAISER FAM. FOUND., <http://kff.org/medicaid/issue-brief/what-coverage-and-financing-at-risk-under-repeal-of-aca-medicare-expansion/> [<https://perma.cc/WCU8-MDM4>].

³⁸ See Ben Moscovitch et al., *Time to Fix the Black Hole in Medicare Data*, HEALTH AFF. BLOG (June 29, 2016), <http://healthaffairs.org/blog/2016/06/29/time-to-fix-the-black-hole-in-medicare-data/> [<https://perma.cc/2LTR-JTB3>] (discussing CMS’s new initiative giving patients, manufacturers, and researchers access to information from health insurance claims forms).

research,³⁹ access has been limited. Files with patient identifiers exclude names and social security numbers,⁴⁰ and numerous security measures are in place.⁴¹

But CMS has opened up data access—from ten datasets that were publicly available in 2009 to more than 2100 datasets in 2016.⁴² Likewise, where data release was previously discretionary, the Affordable Care Act now *mandates* such data reporting to the newly created Patient Centered Outcomes Research Institute to improve clinical and health quality outcomes.⁴³ The Institute has shown a keen interest in secondary research and data collection from patients and provider networks, including clinical, genetic, social/behavioral, and demographic information.⁴⁴ However, the only guaranteed, mandated source of data comes from CMS. Once CMS's EHR data collection initiative is completed, it will be a boon for the Institute's programs. The Institute, in turn, is required to disseminate its research findings for general public use.⁴⁵ Recent legislation and policy have further

³⁹ The data is made available under 42 U.S.C. § 1306(a)–(b) (2012), which authorizes the Department of Health and Human Services (HHS) to release the data pursuant to regulation. The Privacy Act of 1974 requires statements of routine uses of the data in the system to be published. 5 U.S.C. § 552a(e)(4)(D) (2012). HHS has issued the appropriate statements with respect to its various data repositories. One of the routine uses includes “facilitat[ing] research on the quality and effectiveness of care provided.” Privacy Act of 1974; Report of a Modified or Altered System of Records, 71 Fed. Reg. 67,137, 67,137 (proposed Nov. 20, 2006); *see also* *Systems of Records*, CMS.GOV, <https://www.cms.gov/Regulations-and-Guidance/Guidance/PrivacyActSystemofRecords/Systems-of-Records.html#> [<https://perma.cc/N443-KG2H>] (last updated Mar. 2, 2013) (providing a list of “CMS's Systems of Records (SOR) as published in the Federal Register”).

⁴⁰ *See* *Difference Between RIF, LDS, and PUF Data Files*, RESDAC, <https://www.resdac.org/resconnect/articles/148> [<https://perma.cc/3PSU-MTC4>] (last updated Aug. 10, 2016) (explaining that files categorized as “Research identifiable files” (RIFs) exclude certain information so as to protect the individual's privacy). For an example of a state program that also makes Medicaid data available, *see* *Accessing DHCS Protected Data for Research and Public Health*, CAL. DEPT HEALTH CARE SERVICES, <http://www.dhcs.ca.gov/dataandstats/data/pages/accessingprotecteddata.aspx> [<https://perma.cc/CM8F-EPYD>] (last updated Mar. 15, 2017).

⁴¹ *See* *Research Identifiable Files (RIF) Requests*, RESDAC, <https://www.resdac.org/cms-data/request/research-identifiable-files> [<https://perma.cc/7PSE-BFG2>] (noting that CMS's Privacy Board reviews all requests for RIFs).

⁴² Susannah Fox, *Health Datapalooza: New Vistas*, HHS IDEA LAB, (May 20, 2016), <http://www.hhs.gov/blog/2016/05/20/health-datapalooza-new-vistas.html> [<https://perma.cc/9XTW-8RLU>].

⁴³ *See* 42 U.S.C. § 1320e(d)(3)(A) (“The Secretary shall . . . make available to the Institute such data collected by the Centers for Medicare & Medicaid Services . . .”). *See infra* text accompanying notes 263–68 for further discussion of the Institute.

⁴⁴ In 2014, the Institute developed PCORnet, which consists of thirty-three large private clinical networks or health systems and patient networks that have already built themselves to existing capacity. The grants it distributes to enhance collection and coordination are often awarded only where there is already existing stakeholder support and technological infrastructure. *About PCORnet*, PCORNET, <http://www.pcornet.org/about-pcornet/> [<https://perma.cc/9VM5-C3KL>] (last updated Mar. 7, 2017).

⁴⁵ *See* 42 U.S.C. § 1320e(d)(8)(A) (“The Institute shall, not later than 90 days after the conduct or receipt of research findings under this part, make such research findings available to clinicians, patients, and the general public.”).

spurred dissemination. As of last year, CMS has started releasing identifiable data to commercial companies under its new “Innovator” program.⁴⁶

The ACA and subsequent 2015 legislation expand access by limiting discretion and requiring CMS to give individual Medicare data to additional qualified entities.⁴⁷ Under the 2015 program, CMS has stated it will release 100% of beneficiary data provided the qualified entity is able to leverage and show use for it.⁴⁸ This data also includes names and social security information.⁴⁹ The program incentivizes greater amounts of data transmission than first meet the eye: qualified entities will only receive data if they have obtained data from other sources.⁵⁰ They may also constitute a network of organizations rather than a single organization.⁵¹ Entities can access the data for as long as they are part of the program.⁵² The qualified entities may release or sell information, including patient names, to suppliers that have a treatment relationship with the patient or to suppliers that dispute any

⁴⁶ See Telephone Interview with Waruiru Mburu, Tech. Advisor, ResDAC, (July 6, 2016); see also *Innovator Research*, RESDAC, <https://www.resdac.org/cms-data/request/innovator-research> [<https://perma.cc/C7LH-3KBX>] (“Innovators and entrepreneurs may now access CMS data as part of a research study to create products or tools that they intend to sell or to conduct research that creates analyses related to their own business needs.”). Until last year, only limited data sets would be released to these companies. Telephone Interview, *supra*. Admittedly, unlike other data transmissions, the Innovator program requires companies to analyze the data using a CMS portal. *Id.*

⁴⁷ Medicare Access and CHIP Reauthorization Act of 2015, Pub. L. No. 114-10, § 105, 129 Stat. 87, 133. Most of these changes were put in place to help reform payment methodologies. Indeed, the Medicare and CHIP Reauthorization Act (MACRA) has been referred to as possibly “the second most important law to reform the United States’ health care system after the Social Security Amendments which created Medicare and Medicaid in 1965.” Niam Yaraghi, *MACRA Proposed Rule Creates More Problems than It Solves*, HEALTH AFF. BLOG (Oct. 12, 2016), <http://healthaffairs.org/blog/2016/10/12/macra-proposed-rule-creates-more-problems-than-it-solves/> [<https://perma.cc/7JBS-RZJY>].

⁴⁸ Medicare Program; Availability of Medicare Data for Performance Measurement, 76 Fed. Reg. 76,542, 76,552 (Dec. 7, 2011).

⁴⁹ See *id.* at 76,551 (explaining that all claims provided to qualified entities will contain a unique and encrypted beneficiary identification number). Encryption, it should be noted, consists only of removing “direct identifi[er]s” like name, address, and phone number. *Id.* at 76,567; see also 42 C.F.R. § 401.703(f)–(g) (2012) (describing the same).

⁵⁰ Medicare Program; Availability of Medicare Data for Performance Measurement, 76 Fed. Reg. at 76,545–46. CMS has declined to provide a fixed threshold on how much outside data is desirable. Instead, qualified entities are incentivized to maximize data collection. Medicare Program: Expanding Uses of Medicare Data by Qualified Entities, 81 Fed. Reg. 5397, 5399 (proposed Feb. 2, 2016). The recent regulation now seeks to “encourage issuers to submit data . . . to . . . increas[e] . . . sample size.” *Id.*

⁵¹ Medicare Program; Availability of Medicare Data for Performance Measurement, 76 Fed. Reg. at 76,543. Further, CMS declined to limit qualified entity status to nonprofits and governmental bodies, opening it up to commercial organizations as well. *Id.* at 76,544.

⁵² See *id.* at 76,559 (requiring that “once an entity voluntarily leaves or is involuntarily terminated from the program it must destroy or return all CMS data provided . . . within 30 days”). However, they can reapply every three years. *Id.* at 76,568.

performance metric that the qualified entity calculates.⁵³ CMS encourages this kind of data sharing.⁵⁴

Because only claims data is currently collected, the nature of the research projects are limited. However, as its EHR data collection program is completed, CMS data sources will drive additional forms of clinical research that now use private data.⁵⁵ Indeed, the National Academies have called on CMS to “identify or develop and validate measurement standards for collection of [yet even] new social risk factors.”⁵⁶

B. *The Privately Insured*

Private data collection involves a great variety of configurations. Private networks have proprietary databases managed by research subdivisions or outside research entities.⁵⁷ In spite of innovative uses of data in private networks, there are three key differences from CMS data policies. First, although it is difficult to determine how much proprietary research private entities carry out, anecdotes and conversations suggest that many, if not most, private entities do not even engage in internal research. Rather, they use the

⁵³ See Medicare Program: Expanding Uses of Medicare Data by Qualified Entities, 81 Fed. Reg. at 5415 (allowing disclosure regarding information about a supplier’s patients). However, CMS also explains that before releasing either a public or nonpublic analysis, a qualified entity must inform providers and suppliers of information. *Id.* at 5399. Suppliers might seek any identifiable claims data (including names) that are “relevant to the particular measure or measure result” in dispute. 42 C.F.R. § 401.717(c) (2013). This might potentially permit release of non-patient names if non-patient information went into creating the particular measure.

⁵⁴ See Medicare Program; Availability of Medicare Data for Performance Measurement, 76 Fed. Reg. at 76,558 (“We encouraged qualified entities to share this data with providers or supplies upon request.”).

⁵⁵ Part of the problem is that many entities were unable to determine where to transmit data. However, in September 2016, CMS confirmed that it is continuing the task of building a list of public health agencies and clinical data registries that can receive electronic public health measures under new meaningful use reporting requirements. See *CMS to Develop Data Repository for Clinical Data Registry Reporting*, HIT CONSULTANT (Sept. 20, 2016), <http://hitconsultant.net/2016/09/20/cms-to-develop-clinical-data-registry/> [<https://perma.cc/J6NK-X2T4>] (announcing CMS plans to develop a centralized data repository to support sharing public health data).

⁵⁶ THE NAT’L ACADS. OF SCIS., ENG’G., & MED., ACCOUNTING FOR SOCIAL RISK FACTORS IN MEDICARE PAYMENT 26 (2016). CMS is regularly asked to release additional data. For example, last year, House Ways and Means Committee leaders of both parties called on CMS to release more data regarding the mental and behavioral health of Medicare patients in keeping with CMS’s recent open data efforts, such as data on anxiety disorders, bipolar disorder, personality disorders, and traumatic brain injury. See Letter from Representative Kevin Brady et al., to Andrew Slavitt, Acting Adm’r., CMS (Apr. 21, 2016), <http://goo.gl/KgEiAL> [<https://perma.cc/EQ47-MB52>] [hereinafter Brady Letter].

⁵⁷ See, e.g., *Epidemiology and Health Services Research Center*, GEISINGER CARING, <https://www.geisinger.edu/en/research/departments-and-centers/epidemiology-and-health-services-research-center> [<https://perma.cc/3VW2-SWEA>] (providing an example of such a subdivision).

data to provide continuous care as patients circulate among various providers in the network.

Second, even if data is used for internal research to benefit plan members, the research is not generally made publicly available. As the Director for Informatics Research at a prominent medical system explained to me in a confidential email, CMS makes its data public because “[t]hey need to show” Congress and others “that they are doing good things, which is why they . . . are making the data available.” But private entities are known for seeking to maintain data monopolies.⁵⁸ As my source explains, “The data has a certain amount of inherent wealth. The recipient of the data will . . . benefit from using [or] publishing on the data. If something goes wrong then [private entities] (and to a certain extent the recipient [or] secondary user of the data) will be held accountable.” Thus, there is no reason to “give the data to a competitor who gets the accolades, the grants, [or] a competitive advantage.”⁵⁹ As two prominent research scientists argue, this results in private, siloed “war chests of data,” which these entities “enter . . . into systems that are already optimized (primarily for advertising) to make predictions about individuals.”⁶⁰ This makes it harder to agglomerate private data for clinical research.

Finally, in the rare cases in which data is passed on to collaborators and outsiders, unlike CMS data, the data is always stripped of all key identifiers.⁶¹ This means that the risk of identification is lessened.

To be sure, until last year, efforts were well underway in some states to ensure access to the claims information of the privately insured. Almost twenty states have all-payer claims databases (APCDs) in place, which require insurers to send data to central state databases for further analysis.⁶² Such databases have historically been the largest source of private claims data

⁵⁸ For example, Myriad has the breast cancer gene sequence for thousands of variants, but unlike other laboratories, it refuses to share it with ClinVar, the government database of disease-related gene sequences. Erika Check Hayden, *Myriad Genetics Embroiled in Breast-Cancer Data Fight—Again*, NATURE (May 20, 2016), <http://www.nature.com/news/myriad-genetics-embroiled-in-breast-cancer-data-fight-again-1.19953> [<https://perma.cc/V22K-ANZL>].

⁵⁹ *Id.* An example of this is athenahealth’s attempts to address the Zika virus. Athena advertised how it tracked medical records to track patients who may have the virus. Press Release, Athenahealth, Athenahealth Partners with Affected Florida Community to Combat Zika Virus (Aug. 3, 2016), <http://newsroom.athenahealth.com/phoenix.zhtml?c=253091&p=irol-newsArticle&ID=2192379> [<https://perma.cc/Z9YQ-DGL7>]. But it never actually shared the data with public entities or made it available for research—that would diminish its advantage.

⁶⁰ John T. Wilbanks & Eric J. Topol, *Stop the Privatization of Health Data*, NATURE (July 29, 2016), <http://www.nature.com/news/stop-the-privatization-of-health-data-1.20268> [<https://perma.cc/NW2A-LXJF>].

⁶¹ As mandated by the Health Insurance Portability and Accountability Act of 1996 (HIPAA), data is generally only shared if eighteen identifiers are stripped, but limited data sets can be shared with appropriate agreements in place. *See* 42 C.F.R. § 164.514(b)(2)(i)(A)–(R) (2011) (listing the types of identifying information that must be removed under HIPAA). Moreover, entities can hire a data scientist to certify that the risk of reidentification is minimal. *Id.* § 164.514(b)(1).

⁶² *Gobeille v. Liberty Mut. Ins. Co.*, 136 S. Ct. 936, 941 (2016).

for research.⁶³ States used these databases for research, but also sold de-identified data for commercial purposes to fund these all-payer efforts.⁶⁴

However, in 2016, the Supreme Court held that the Employee Retirement Income Security Act of 1974 (ERISA) preempted states from mandatorily collecting claims data from self-funded plans.⁶⁵ With this decision, the Court “bl[e]w an enormous hole in the all-payer claims databases” as two-thirds of all employees are self-insured.⁶⁶ In July, the Department of Labor issued proposals for information collection, but “it is hard to see that the information [it] proposes to collect . . . would be an adequate substitute for the all-payer claims database information.”⁶⁷

C. High-Income Individuals

Finally, those who are wealthier, healthier, or younger—characteristics which often go together⁶⁸—are less likely to have health data in the publicly

⁶³ See Sean E. Bland et al., *Strategies for Health System Innovation After Gobeille v Liberty Mutual Insurance Company*, 316 J. AM. MED. ASS'N. 581, 581 (2016) (“Over the past decade, many states have used APCDs to capture new and more sophisticated insurance claims and utilization data . . .”).

⁶⁴ See *Resources: Frequently Asked Questions*, APCD COUNCIL, <https://www.apcdouncil.org/frequently-asked-questions> [https://perma.cc/23M6-C6TV] (“Agencies with laws and policies that permit the release of standard de-identified and research APCD analytic files can generate revenues from the sales of these products, with the appropriate release agreements and research review approvals.”).

⁶⁵ *Gobeille*, 136 S. Ct. at 947.

⁶⁶ Nicholas Bagley, *The Supreme Court's Wrongheaded Decision in Gobeille*, INCIDENTAL ECONOMIST (Mar. 3, 2016, 11:50 AM), <http://theincidentaleconomist.com/wordpress/the-supreme-courts-wrongheaded-decision-in-gobeille/> [https://perma.cc/JKH5-LUP2]. This is the majority view. See e.g., Bland et al., *supra* note 63 (arguing that the *Gobeille* decision “may undermine opportunities for health system innovation” and proposing solutions to address the problem). But a minority view suggests that the decision may not be so fatal. See David Newman et al., *Losing the All' In All-Payer Claims Databases*, HEALTH AFF. BLOG (July 18, 2016), <http://healthaffairs.org/blog/2016/07/18/losing-the-all-in-all-payer-claims-databases/> [https://perma.cc/44WR-S6Z4] (“While we understand that the potential loss of ERISA data is viewed with concern . . . , the Court's decision may not be fatal to policy-relevant research.”).

⁶⁷ Tim Jost, *Labor, IRS Propose New Health Plan Reporting Requirements; CMS Makes Its Case on Cost Sharing* (July 12, 2016), <http://healthaffairs.org/blog/2016/07/12/labor-irs-propose-new-health-plan-reporting-requirements-cms-makes-its-case-on-cost-sharing/> [https://perma.cc/WLZ8-KWUU]; see also Annual Reporting and Disclosure, 81 Fed. Reg. 47,496, 47,499-47,500 (proposed July 21, 2016) (to be codified at 29 C.F.R. pts. 2520, 2590) (proposing to expand reporting requirements by “eliminating obsolete exemptions for certain plans”). In defense of Labor, while they sought comment on how new data collection requirements related to *Gobeille*, the data collection was likely planned before *Gobeille*, and thus sought only to satisfy Labor's specific mandate under the Affordable Care Act for quality improvement efforts of health plans. *Id.* at 47,528 (citing 29 U.S.C. § 1185(d) (2012)).

⁶⁸ Young people tend to be poorer than rich people. But young people who are healthier tend to be wealthier, and healthier people, holding age constant, tend to be rich. See STEVEN H. WOLF ET AL., URBAN INST., & CTR. ON SOC'Y & HEALTH, HOW ARE INCOME AND WEALTH LINKED TO HEALTH AND LONGEVITY? 1 fig.1 (2015), <http://www.urban.org/sites/default/files/alfresco/publication-pdfs/2000178-How-are-Income-and-Wealth-Linked-to-Health-and-Longevity.pdf>

available information pot. As of September 2013, pursuant to regulatory changes, a provider *must* comply with a patient's request not to disclose information to health plans if the patient pays for the service out of pocket.⁶⁹

Although the law contemplates that patients can prevent doctors from disclosing data only to health plans, there is reason to think that some providers could simply omit the data from the health record where possible. In its overview of the comments to the 2013 rule, the Department of Health and Human Services (HHS) acknowledged that providers found the new requirement burdensome as they would “hav[e] to . . . manually redact information from the medical record” or even “create separate records” prior to an audit—when an insurer can access most provider files.⁷⁰ HHS provided no real guidance to address this concern, suggesting only that providers adopt data minimization techniques.⁷¹ Thus, where providers are not required to retain information by law or treatment standards, the simplest way to avoid having to segregate records to guard against inadvertent disclosure and HIPAA liability is to omit recording information altogether.

But the truly rich do not need the benefit of HIPAA rules. Many of the richest individuals rely on “concierge medicine.” According to a recent profile, one such practice has a highly limited number of patients, each of whom pays \$25,000 a year

for unfettered access to the doctors. Patients will be able to call and see and text the doctors whenever they want; they will be able to receive home visits They will be able to ask their doctors to travel to them should they suspect the onset of illness in June in Umbria.⁷²

[<https://perma.cc/S7HB-EAZY>] (tracking the likelihood of death and disease relative to economic standing).

⁶⁹ 45 C.F.R. § 164.522(a)(1)(vi) (2013); *see also* Modifications to the HIPAA Privacy, Security, Enforcement, and Breach Notification Rules Under the Health Information Technology for Economic and Clinical Health Act and the Genetic Information Nondiscrimination Act; Other Modifications to the HIPAA Rules, 78 Fed. Reg. 5566, 5624 (Jan. 25, 2013) (adopting a final rule that “restrict[s] certain disclosures of protected health information to a health plan where the individual pays out of pocket in full for the health care item or service”).

⁷⁰ Modifications to the HIPAA Privacy, Security, Enforcement, and Breach Notification Rules Under the Health Information Technology for Economic and Clinical Health Act and the Genetic Information Nondiscrimination Act; Other Modifications to the HIPAA Rules, 78 Fed. Reg. at 5627.

⁷¹ *See id.* at 5628 (explaining that providers should already “have in place . . . minimum necessary policies and procedures, which require limiting the protected health information disclosed to a health plan”).

⁷² Ginia Bellafante, *Enhanced Medical Care for an Annual Fee*, N.Y. TIMES (Dec. 6, 2013), http://www.nytimes.com/2013/12/08/nyregion/enhanced-medical-care-for-an-annual-fee.html?_r=0 [<https://perma.cc/A26Y-HZ72>].

As of 2008, there were twenty-eight concierge doctors in the New York metropolitan area, and by 2013, there were 124.⁷³ That number may continue to grow.⁷⁴

Finally, because nearly all health information comes from providers, those who do not need to go to providers remain completely insulated from the collection system. These individuals skew comparatively healthy, wealthy, and young.⁷⁵ They can rely on wearable devices like FitBits or, increasingly, wellness programs, to monitor their health without ever seeing a provider.⁷⁶ Nonetheless, their data is necessary to develop controls for research involving illness, including information about normal variation,⁷⁷ and asymptomatic versions of certain conditions.⁷⁸

II. THE BURDENS OF HEALTH INFORMATION COLLECTION

Understanding the need for equity in health information collection requires appreciating its burdens. Health information collection, storage, and

⁷³ *Id.*

⁷⁴ *See id.* (“The risk of course is that these sort of practices multiply . . .”).

⁷⁵ *See* CTRS. FOR DISEASE CONTROL & PREVENTION, SUMMARY HEALTH STATISTICS: NATIONAL HEALTH INTERVIEW SURVEY, 2014 tbl.A-18a (undated), http://ftp.cdc.gov/pub/Health_Statistics/NCHS/NHIS/SHS/2014_SHS_Table_A-18.pdf [<https://perma.cc/MY7F-MSL8>] [hereinafter CDC STATISTICS]. As the data reveals, younger individuals tend to have less contact with medical professionals. *Id.* Holding age constant, individuals with private insurance visit the doctor less frequently than CMS populations. *Id.* This is likely—at least in part—due to the fact that they are healthier. *See* Paradise & Garfield, *supra* note 31 (“Medicaid beneficiaries are poorer and have a poorer health profile compared with both the privately insured and the uninsured.”).

⁷⁶ *See* Nick Bilton, *Where Wearable Technology Ends Up (Hint: Not Your Wrist)*, N.Y. TIMES (Jan. 20, 2016), http://www.nytimes.com/2016/01/21/style/where-wearable-technology-ends-up-hint-not-your-wrist.html?_r=0 [<https://perma.cc/D42E-W8A4>] (describing the wearable fitness tracker industry as a multi-billion dollar industry, albeit one in decline).

⁷⁷ *See* Adeline R. Whitney et al., *Individuality and Variation in Gene Expression Patterns in Human Blood*, 100 PROC. NAT’L ACAD. OF SCI. 1896, 1900-01 (2003) (using data from healthy individuals to determine normal variation in gene expression). In fact, a 2012 survey of “6200 doctors, nurses, lay people, and legislators in Finland found wide disagreement over whether dozens of so-called ‘states of being’ should even be considered diseases, including ADHD and overactive bladder.” John Fauber et al., *Lowering the Bar: Medicine in the 21st Century*, MEDPAGE TODAY (May 22, 2016), <http://www.medpage.com/special-reports/LoweringtheBar/58048> [<https://perma.cc/6QAH-SQXD>].

⁷⁸ Admittedly data from sicker patients is still thought to be more useful. Thus, records of sick patients contain more information than those of healthy patients. *See* Nicole G. Weiskopf et al., *Sick Patients Have More Data: The Non-Random Completeness of Electronic Health Records*, 2013 AMIA ANN. SYMP. PROC. 1472, 1476 (“Sicker patients tend to have more complete records and healthier patients tend to have records that are less complete.”). But healthy individuals’ information is increasingly being used for various purposes, including—for example, trying to understand what sickness itself is. *See* Hans K. Meier-Ewert et al., *Absence of Diurnal Variation of C-Reactive Protein Concentrations in Healthy Human Subjects*, 47 CLINICAL CHEMISTRY 426, 426 (2001) (noting that healthy subjects’ data can be used to predict future risk of myocardial infarction and stroke). *See generally* Carol Masheter et al., *Making Cents of Utah’s Healthy Population*, 1 UTAH ATLAS HEALTHCARE, Oct. 2010, at 1 (explaining Utah’s effort to collect information about healthy individuals).

use impose burdens—both material and autonomy based—on individuals.⁷⁹ This Part describes a range of burdens that an individual might suffer by having her data used. But imposing burdens is not, in itself, normatively unacceptable—we justify doing so in a range of cases. Thus, this Part is primarily descriptive in nature.

A. Privacy and Security Harms

Each act of data transmission presents some risk that data will be breached, even with security precautions in place. The frequency of breaches has increased.⁸⁰ In 2015, 113 million electronic health records were breached.⁸¹ Almost ninety percent of health care covered entities admitted to a patient data breach in the last two years.⁸²

Criminal attacks were the leading cause of these breaches, accounting for about half of them.⁸³ Health care institutions are more vulnerable to these attacks.⁸⁴ Hospital “hostage taking” has frequently made the news recently, where hackers have retained hospital data until they were paid sums of money.⁸⁵

⁷⁹ This distinction between material burdens and autonomy-related burdens tracks the distinction in Calo, *supra* note 7, at 1133, between objective and subjective burdens.

⁸⁰ See PONEMON INST., SIXTH ANNUAL BENCHMARK STUDY ON PRIVACY & SECURITY OF HEALTHCARE DATA 1 (2016) (discussing “the increased frequency of [data] breaches” in the health care industry).

⁸¹ U.S. GOV’T ACCOUNTABILITY OFF., GAO-16-771, REPORT TO THE COMMITTEE ON HEALTH EDUCATION, LABOR, AND PENSIONS, U.S. SENATE, ELECTRONIC HEALTH INFORMATION: HHS NEEDS TO STRENGTHEN SECURITY AND PRIVACY GUIDANCE AND OVERSIGHT 9 (2016), <http://www.gao.gov/assets/680/679260.pdf> [<https://perma.cc/7DCZ-SRAN>]; see also *id.* at 13 (“[T]he magnitude of the threat against health care information has grown exponentially.”). Another group predicted that one in three health care recipients would be the victim of a medical data breach in 2016. Dan Munro, *Data Breaches in Healthcare Totaled over 112 Million Records in 2015*, FORBES (Dec. 31, 2015, 9:11 PM), <http://www.forbes.com/sites/danmunro/2015/12/31/data-breaches-in-healthcare-total-over-112-million-records-in-2015/#7c89ac677fd5> [<https://perma.cc/2AJN-8SDK>].

⁸² PONEMON INST., *supra* note 80, at 1.

⁸³ *Id.*

⁸⁴ *Id.* at 2. In fact, the health care industry led all industries with eighty-eight percent of all ransomware detected in the second quarter of 2016. *SERT Quarterly Threat Report Q2 2016*, NTT SECURITY (June 26, 2016), <https://www.solutionary.com/threat-intelligence/threat-reports/quarterly-threat-reports/sert-threat-report-q2-2016/> [<https://perma.cc/95JP-RECR>]. And a Brookings Institution study predicts that one in four data breaches this year will hit the health care industry. NIAM YARAGHI, BROOKINGS INST., HACKERS, PHISHERS, AND DISAPPEARING THUMB DRIVES: LESSONS LEARNED FROM MAJOR HEALTH CARE DATA BREACHES 1 (2016) (“Twenty-three percent of all data breaches happen in the health care industry.”). The FBI, similarly, anticipates rising cyber threats to health care. See Joseph Goedert, *FBI Sees Rising Cyber Threats to Healthcare*, HEALTH DATA MGMT. (July 13, 2016, 7:11 AM), <http://www.healthdatamanagement.com/news/fbi-sees-rising-cyber-threats-to-healthcare-organizations> [<https://perma.cc/386V-Y7KJ>] (“The Federal Bureau of Investigation sees increasing presence from hackers trying to access patient information from providers.”).

⁸⁵ See e.g., Jeff Stone, *Ransomware Hackers a Bigger Threat than Ever, Forcing Hospitals and Police to Pay Hostage Fees*, INT’L BUS. TIMES (Feb. 23, 2016, 2:46 PM), <http://www.ibtimes.com/>

This has led to discussions among members of the cybersecurity community and in Congress to increase hospital data security standards.⁸⁶ Health care institutions report that they are the least prepared to deal with these breaches compared to entities in other sectors.⁸⁷ As the scope and profitability of secondary research increases, the focus of hackers will likely shift to health data repositories.⁸⁸

Further, as one prominent government regulator notes, “[L]egislating security is really tricky You write down requirements and the bad guys immediately supersede them.”⁸⁹ Many of the breaches or security compromises involve government entities⁹⁰ (although there is no evidence that government entities are, on average, less secure than private entities).

ransomware-hackers-bigger-threat-ever-forcing-hospitals-police-pay-hostage-fees-2319822 [https://perma.cc/ED5G-NZ9J] (describing a “cyberattack that left a California hospital paralyzed until administrators agreed to pay a \$17,000 ransom”).

⁸⁶ *Id.*

⁸⁷ According to a survey, “An overwhelming majority of health care organizations (69 percent) and business associates (63 percent) believe they are at greater risk than other industries for a data breach.” PONEMON INST., *supra* note 80, at 3. Only 29 percent in health care said they are prepared to handle external cyber threats, far lower than any other industry. Greg Slabodkin, *Healthcare and Pharma Least Prepared for External Cyber Threats*, HEALTH DATA MGMT. (July 18, 2016, 3:33 PM), <https://www.healthdatamanagement.com/news/healthcare-and-pharma-least-prepared-for-external-cyber-threats> [https://perma.cc/Y4UD-YGV7]. Only a quarter in health care feel they can analyze the threats, which also lags other industries. *Id.*

⁸⁸ See *Are EHR Vendors Hackers' Next Big Target?*, HIT CONSULTANT (Apr. 11, 2016), <http://hitconsultant.net/2016/04/11/preparing-ehr-vendors-cyber-threats/> [https://perma.cc/D87K-9Q6G] (predicting that electronic health record providers will be increasingly targeted); see also ERNST & YOUNG, *NAVIGATING THE BULL BLACK MARKET* 7-9 (2016), [https://www.ey.com/Publication/vwLUAssets/ey-navigating-the-bull-black-market/\\$FILE/ey-navigating-the-bull-black-market.pdf](https://www.ey.com/Publication/vwLUAssets/ey-navigating-the-bull-black-market/$FILE/ey-navigating-the-bull-black-market.pdf) [https://perma.cc/KY3B-XRRW] (noting that pharmaceutical and life sciences firms will be targeted for clinical trials data).

⁸⁹ David Pittman, *AMA Meeting Under Way in Chicago*, POLITICO: MORNING EHEALTH (June 13, 2016, 10:00 AM), <http://www.politico.com/tipsheets/morning-ehealth/2016/06/politicos-morning-ehealth-ama-meeting-under-way-in-chicago-hipaa-confusion-in-orlando-shooting-state-seeks-cloud-based-ehr-214786> [https://perma.cc/788X-B7BP] (internal quotation marks omitted).

⁹⁰ In a two month period alone, government audits have pointed to “numerous weaknesses” in the security of Minnesota’s health insurance exchange. AMY J. FRONTZ, OFFICE OF THE INSPECTOR GEN., DEP’T OF HEALTH & HUMAN SERVS., *PUBLIC SUMMARY REPORT: INFORMATION TECHNOLOGY CONTROL WEAKNESSES FOUND AT THE MINNESOTA HEALTH INSURANCE EXCHANGE 2* (2016). In addition, audits have found evidence of weak data encryption, too many staff members with access to sensitive data, and inconsistent auditing of network activity in the FDA. U.S. GOV’T ACCOUNTABILITY OFF., GAO-16-513, *INFORMATION SECURITY: FDA NEEDS TO RECTIFY CONTROL WEAKNESSES THAT PLACE INDUSTRY AND PUBLIC HEALTH DATA AT RISK* 11-19 (2016). Furthermore, such audits have identified many problems and conflicts facing federal agencies’ Chief Information Security Officers. U.S. GOV’T ACCOUNTABILITY OFF., GAO-16-686, *FEDERAL CHIEF INFORMATION SECURITY OFFICERS: OPPORTUNITIES EXIST TO IMPROVE ROLES AND ADDRESS CHALLENGES TO AUTHORITY* 26-31 (2016). One recent hack involved a contractor of the Milwaukee Veterans Health Affairs Medical Center. See *Milwaukee Veterans Health Information Hacked*, WSAW-TV (Sept. 2, 2016, 11:31 PM), <http://www.wsaw.com/content/news/Milwaukee-veterans-health-information-hacked-392226671.html> [https://perma.cc/K69W-CWTE]

Data breaches have led to instances of identity theft, medical and financial fraud, tainted medical records, and insurance discrimination. Scholars have also pointed to the risk of employment discrimination that data breaches pose.⁹¹ In another instance from last summer, a hacker put medical records up for sale on the Internet.⁹² But, as the Brookings Institution observes,

(describing the hacking of a Wisconsin Medical College employee's email). I focus here on data collection and transmission for secondary research by the government. But the few others to comment on these data policies are troubled by the initial transmission of data to the government. For example, language in another proposed rule states that a "health care provider is permitted to disclose protected health information (PHI) (without patient authorization and without a business associate agreement)" to a government entity so that it can "perform the required on-site surveillance of the certified EHR technology." Medicare Program; Merit-Based Incentive Payment System (MIPS) and Alternative Payment Model (APM) Incentive Under the Physician Fee Schedule, and Criteria for Physician-Focused Payment Models, 81 Fed. Reg. 28,162, 28,171 (proposed May 9, 2016) (to be codified at 42 C.F.Rs. pts. 414, 495). One prominent commentator from the Brookings Institution argues that such surveillance "is extremely dangerous and inefficient." Yaraghi, *supra* note 47. I agree that such surveillance is unnecessary in this case, but in the case of secondary research, data collection is necessary. But those burdens should be imposed in a fair manner.

⁹¹ Health data breaches often involve names and social security numbers, such as in a recent breach involving Massachusetts General Hospital (MGH) data. See Jessica Bartlett, *MGH Says Patients Impacted by Third-Party Data Breach*, BOS. BUS. J. (June 29, 2016, 4:43 PM), <http://www.bizjournals.com/boston/blog/health-care/2016/06/mgh-says-patients-impacted-by-third-party-data.html> [<https://perma.cc/MZ36-9AG9>] (describing a 2016 hack of MGH's software provider, which exposed patient names, birthdays, social security numbers, and medical research). This results in identity theft and financial and medical fraud. See Jessica Davis, *Ransomware and Tax Fraud Rise as Healthcare Hit with More Data Breaches than Any Other Industry*, HEALTHCARE IT NEWS (April 13, 2016, 7:58 AM), <http://www.healthcareitnews.com/news/ransomware-and-tax-fraud-rise-healthcare-hit-more-data-breaches-any-other-industry> [<https://perma.cc/GD2N-QL4A>] (cataloguing the rise of tax fraud and the issues faced by victims); IBM, *70% of Businesses Paid Cybercriminals to Unlock Ransomware*, HIPAA J. (Dec. 15, 2015), <http://www.hipaajournal.com/phi-theft-by-employees-peacehealth-8210/> [<https://perma.cc/CD5N-U5TV>] (describing instances of medical record theft). This may also muddy medical histories. See Darius Tahir & Bob Herman, *Data Breaches Can Lead to Major Medical Identity Theft Issues*, MOD. HEALTHCARE (Mar. 4, 2015), <http://www.modernhealthcare.com/article/20150304/NEWS/150309960> [<https://perma.cc/TB6K-YCVG>] (noting that healthcare data breaches can result in a "completely altered medical history" for patients). Health records can be used for other medical purposes, such as creating fake opioid prescriptions. See Fahmida Y. Rashid, *Why Hackers Want Your Health Care Data Most of All*, INFOWORLD TECH WATCH (Sep. 14, 2015), <http://www.infoworld.com/article/2983634/security/why-hackers-want-your-health-care-data-breaches-most-of-all.html> [<https://perma.cc/W5AL-9WFP>] ("Stolen health care data forums operate more like drug cartels, where health records are not sold outright, but rather used to buy and sell addictive prescriptions."); see also Frank Pasquale, *Redescribing Health Privacy: The Importance of Information Policy*, 14 HOUS. J. HEALTH L. & POL'Y 95, 98 (2014) (identifying insurance fraud as an example of how patient identify records can be misused).

⁹² Dissent Doe, *655,000 Patient Records For Sale on the Dark Net After Hacking Victims Refuse Extortion Demands*, DAILY DOT (June 27, 2016, 7:29 AM), <https://www.dailydot.com/layer8/655000-patient-records-dark-net/> [<https://perma.cc/C7H3-3S9Y>]. The hacker offered to sell the personal information of 655,000 patients whose data were stolen from three databases that refused to pay a ransom. *Id.* Specifically, 48,000 records from a Farmington, Missouri database were offered for \$100,000; 397,000 records from an Atlanta database were offered for \$400,000; and 210,000 records from a Central/Midwest database were offered for \$200,000. *Id.* The data include names, addresses, Social Security numbers, birthdates, emails, gender, and phone numbers. *Id.*

“Despite the public concerns over health care privacy breaches, we do not know exactly why hackers are interested in stealing medical data or how exactly they monetize it.”⁹³ For example, hackers may hold on to breached records for years before using them to commit medical or financial identity theft or blackmail.⁹⁴

Although there is “a great deal of confusion about the value of stolen medical data in the black market,”⁹⁵ we know that criminal elements value health data. Experts uniformly report that health data is more valuable than other data, and that its value is on the rise.⁹⁶ And information professionals, even those that support secondary research, are increasingly worried about data breaches.

Breaches affect private and public institutions alike. But because of the new policies described in Part I, the poorer and older bear greater risk. Data risk increases as the range of uses, the number of entities with access, and the amount and kind of data used increases. Human error is often the most important security hurdle as individuals misplace storage devices or transmit data over unsecured connections or to incorrect recipients.⁹⁷ Using identifiers—especially names and social security numbers—also increases risk.⁹⁸

⁹³ Niam Yaraghi, *To Mitigate Medical Hacks, Identify Incentives for Hackers*, BROOKINGS (Aug. 16, 2016), <https://www.brookings.edu/blog/techtank/2016/08/16/to-mitigate-medical-hacks-identify-incentives-for-hackers/> [<https://perma.cc/4EHN-MHXQ>].

⁹⁴ One IBM Security executive learned that the medical records of his two grade-school-age children were compromised in a health care data breach months after it happened. According to the Boston Globe, “Those identities are essentially worthless for a criminal seeking quick payoff—nobody’s going to give a loan to a preteen. But once those kids reach their 18th birthdays . . . , the leak could come back to haunt their finances. ‘How long is it going to be before that 18-year-old realizes that somebody at another address has established credit in their name?’” Curt Woodward, *Health Files Make for a Juicy Target for Thieves*, BOS. GLOBE (Aug. 7, 2016), http://www.bostonglobe.com/business/2016/08/07/hackers-turn-health-care-where-records-fetch-bigger-bucks/OSi8imOSGPstvwV6DyYhoJ/story.html?s_campaign=bostonglobe:socialflow:twitter [<https://perma.cc/25QZ-K9YY>].

⁹⁵ Yaraghi, *supra* note 93.

⁹⁶ See Goedert, *supra* note 84 (“Recent events suggest that the pressure [from hackers trying to access patient information] may be rising . . .”).

⁹⁷ See Jeff Goldman, *December Data Breach Roundup: Theft Prevention, and More*, ESECURITY PLANET (Jan. 21, 2014), <http://www.esecurityplanet.com/network-security/december-data-breach-roundup-theft-prevention-and-more.html> [<https://perma.cc/6XRY-DTGM>] (advocating for employer education as a prevention tool for data breaches).

⁹⁸ Privacy advocates are even worried about the possibility of re-identification of de-identified data. In a high-profile example of this risk, the former Chief Technology Officer of the Federal Trade Commission, Latanya Sweeney, purchased de-identified insurance records from a state insurance database when she was a graduate student, and she was able to identify those of then-Massachusetts Governor William Weld by combining them with other public data. Daniel Barth-Jones, *The Debate Over ‘Re-Identification’ of Health Information: What Do We Risk?*, HEALTH AFF. BLOG (Aug. 10, 2012), <http://healthaffairs.org/blog/2012/08/10/the-debate-over-re-identification-of-health-information-what-do-we-risk/> [<https://perma.cc/6RAK-NQ4Y>].

The data of those in CMS programs has experienced a marked increase in transmission and accumulation—often in identified formats—for multiple uses.⁹⁹ The data of those in private programs is transmitted and used less, and never transmitted externally in identified formats. Those who self-pay may never have their data recorded, much less used.

B. *Personhood Harms*

Studies suggest that individuals are concerned with harms that go beyond material threats from malicious actors. In many circumstances, individuals object to the collection or use of information without consent, even if there is no threat—including when the data is used only for research. This concern is echoed in recent practices. Within the last two years alone, the federal government retained, but then ultimately decided not to implement, a proposal to require informed consent for research involving biospecimen data,¹⁰⁰ and the British National Health Service has come under fire for a project involving the transmission of anonymous data to Google for research.¹⁰¹

The story of Henrietta Lacks presents a good example of the kind of harm to which I am alluding. Lacks was an African-American woman who received indigent medical care in the 1950s. In the process, her cells with unique properties were harvested without her permission. When this was recently

⁹⁹ Moreover, CMS in particular is not fully compliant with breach notification requirements and does not offer remedies to Medicare patients when there are breaches. See *Investigation Faults Handling of Medicare Patient Data Breaches*, AMEDNEWS.COM (Oct. 29, 2012), <http://www.amednews.com/article/20121029/business/310299965/6/> [<https://perma.cc/ADB8-FRZQ>] (noting that the Office of the Inspector General “found that Medicare wasn’t doing enough to mitigate damages caused when a Medicare patient’s identification is stolen”).

¹⁰⁰ The government first proposed this requirement in its Advanced Notice of Proposed Rulemaking for the Common Rule in 2011, and it retained the requirement in its Notice of Proposed Rulemaking. In explaining why it proposed to retain the consent requirement, the government observed, “[T]here is a growing recognition that many people want to have some degree of control over the circumstances in which an investigator can derive information about them, above and apart from their interest in whether or not that information might be inappropriately disclosed. More specifically, a growing body of literature shows that in general people prefer to have the opportunity to consent (or refuse to consent) to research involving their own biological materials [I]t is not consistent with the majority of the public’s wishes, which reflect legitimate autonomy interests.” Federal Policy for the Protection of Human Subjects, 82 Fed. Reg. 53,933, 53,942, 53,944 (proposed Sept. 8, 2015). Notably, the proposed rulemaking cites an infamous incident involving the use of Havasupai genes for research without consent. *Id.* at 53,943 n.43 (citing National Congress of American Indians, *Havasupai Tribe and the Lawsuit Settlement Aftermath*, NAT’L CONGRESS AM. INDIANS, <http://genetics.ncai.org/case-study/havasupai-Tribe.cfm> [<https://perma.cc/42Z9-AXEX>]). However, after receiving significant pressure from industry and other groups, the government decided not to impose this requirement. Federal Policy for the Protection of Human Subjects, 82 Fed. Reg. 7149, 7165 (Jan. 19, 2017).

¹⁰¹ See *Why Google DeepMind Wants Your Medical Records*, BBC (July 19, 2016), <http://www.bbc.com/news/technology-36783521> [<https://perma.cc/FM7U-4VYV>] (describing the criticism that followed the announcement that Google was given access to the health care data of 1.6 million patients).

discovered, the settlement that Lacks's descendants reached did not prohibit the use of their genetic data. Rather, they received some control over who accesses the data and credit for publications involving the data.¹⁰²

The Lacks story is simply an early instance of the growing “biorights” movement. Last May, the American Civil Liberties Union (ACLU) filed suit on behalf of four plaintiffs against Myriad Genetics,¹⁰³ a company notorious for trying to (unsuccessfully) patent genes predicting breast cancer.¹⁰⁴ Myriad obtains genetic samples from patients and lets them know if they are at risk for certain conditions, but does not provide them access to the coded data. The suit asks the Department of Health and Human Services “to ensure patients have access to all the genetic data a testing company gleans from their samples, not simply whether they have a certain gene or condition.”¹⁰⁵ In this case, the plaintiffs want to make the data available for public research.¹⁰⁶

Studies show that when it comes to data use, consent is key—individuals believe that using data without consent is wrong but are happy to provide consent if asked.¹⁰⁷ Their focus appears to be on the dignitary affront and

¹⁰² Malcolm Ritter, *NIH, Family of Henrietta Lacks Reach Deal on Access to DNA Code*, WASH. POST (Aug. 7, 2013), https://www.washingtonpost.com/national/health-science/nih-family-of-henrietta-lacks-reach-deal-on-access-to-dna-code/2013/08/07/68f3da04-ff8b-11e2-96a8-d3b921c0924a_story.html [https://perma.cc/3U58-UHKU].

¹⁰³ Beth Daley & Ellen Cranley, *'Biorights' Rise: Donors Demand Control of Their Samples*, BOS. GLOBE (Oct. 10, 2016), <https://www.bostonglobe.com/metro/2016/10/09/the-rise-biorights-donors-are-demanding-control-and-sometimes-cash-exchange-for-geneticsamples/jCbaQ2E5t6coQs1kclTMRM/story.html> [https://perma.cc/GN2A-C6RA].

¹⁰⁴ See, e.g., Andrew Pollack, *Myriad Genetics Ending Patent Dispute on Breast Cancer Risk Testing*, N.Y. TIMES (Jan. 27, 2015), <https://www.nytimes.com/2015/01/28/business/myriad-genetics-ending-patent-dispute-on-breast-cancer-risk-testing.html> [https://perma.cc/M9MT-68VX] (“Myriad Genetics has essentially given up trying to stop other companies from offering tests for increased risk of breast cancer . . .”).

¹⁰⁵ Daley & Cranley, *supra* note 103. Admittedly, individuals are happy to provide their data in return for *payment*. DNAsimple, for example, pays for data. *Id.* But while whether companies should be allowed to pay for this data is an open question, it cannot be denied that payment at least gives individuals a sense of control. Patients “feel part of the process when they get compensated.” *Id.* (internal quotation marks omitted); cf. Sonia M. Suter, *Disentangling Privacy from Property: Toward a Deeper Understanding of Genetic Privacy*, 72 GEO. WASH. L. REV. 737, 746 (2004) (“Genetic information is seen as a commodity, disaggregated from the self, rather than something in which we have a dignitary and personhood interest.”).

¹⁰⁶ Other examples exist. For example, Medtronic refuses requests for patients' own heart data. See Jonah Comstock, *Medtronic Launches Connected App for Pacemaker Patients, but Patients Can't See the Data*, MOBIHEALTHNEWS (Nov. 18, 2015), <http://www.mobihealthnews.com/48596/medtronic-launches-connected-app-for-pacemaker-patients-but-patients-cant-see-the-data> [https://perma.cc/V4M3-4QKA] (describing a Medtronic mobile app that permits patients to forward pacemaker data to providers but that does not provide patients with actual access to the data). Indeed, “early players in the game have sequestered information in ‘closed loop’ systems.” John T. Wilbanks & Eric J. Topol, *Stop the Privatization of Health Data*, NATURE (July 29, 2016), <http://www.nature.com/news/stop-the-privatization-of-health-data-1.20268> [https://perma.cc/PV77-ZDDP].

¹⁰⁷ See Daley & Cranley, *supra* note 103 (“About 68% of people are willing to give permission for researchers to use their specimens to be used for any purpose But support drop[s] to 55%

sense of violation that comes from sharing data without consent—rather than privacy or security harms.¹⁰⁸

This Article does not seek to identify the precise dignity interest at stake. Rather, I rely on a rough analogy to another kind of dignitary or personhood harm that has appeared in the property literature. Many years ago, Margaret Radin argued that an individual's property is imbued with their personhood.¹⁰⁹ Property enables the creation and execution of future projects, and populates memory with projects past. It helps us conceptualize the individual as a continuous entity through time.¹¹⁰ This creates an empirically demonstrated psychological link between individuals and their property.¹¹¹

Similarly, rights to the information about one's body, mind, and certain objects grant a person autonomy to choose his or her course in life. In discussing the learning health system, President Obama explained that it helps "empower[] individuals" to monitor and take a more active role in their own health."¹¹² Health data can be used to develop essential life goals related

if patients kn[ow] their sample will be used to 'develop patents and earn profits for commercial companies.'"). And a National Institutes of Health (NIH) survey shows that if asked, a majority of Americans (fifty-four percent) said they would "definitely or probably participate" in the NIH's Precision Medicine Initiative Cohort Program, which requires personal data such as blood samples, genetic information, family medical history, soil and water samples from their home, as well as data on their lifestyle, diet, and exercise. *Survey Shows Broad Support for National Precision Medicine Study* (Aug. 17, 2016), NAT'L INSTITUTES HEALTH, <https://www.nih.gov/news-events/news-releases/survey-shows-broad-support-national-precision-medicine-study> [<https://perma.cc/4BHU-EHP7>]. Among younger people, college-educated people, and those who identified as lesbian, gay, bisexual, or transgender, interest in participation was even higher. *Id.*

¹⁰⁸ One study found that 77% of individuals would feel "violated and [that their] trust in the researchers [would be] betrayed" in such a situation. ALAN F. WESTIN, *HOW THE PUBLIC VIEWS PRIVACY AND HEALTH RESEARCH* 25 (rev. 2008). But only 44% percent feared employment discrimination, and just 33% feared embarrassment "before friends, associates or the public." *Id.* Indeed, only 19% of individuals were willing to have their information used without their consent when their anonymity was assured. *Id.* at 28. Similarly, Westin reviewed previous studies that showed a consistent majority who argued that data can only be used with consent. *Id.* at 45-46; see also WESTAT, AHRQ PUB. NO. 09-0081-EF, FINAL REPORT: CONSUMER ENGAGEMENT IN DEVELOPING ELECTRONIC HEALTH INFORMATION SYSTEMS 2 (2009) ("A substantial proportion [of focus group participants] felt that health care consumers owned their data and needed a role in ensuring that those data were secure and used only in ways that they authorized."); Evette J. Ludman et al., *Glad You Asked: Participants' Opinions of Re-Consent to dbGAP Data Submission*, 5 J. EMPIRICAL RES. ON HUM. RES. ETHICS, Sept. 2010, at 9, 13 ("It was very important or somewhat important to the majority . . . of respondents that they were asked for their permission to add their health and genetic information to the databank.").

¹⁰⁹ Margaret Jane Radin, *Property and Personhood*, 34 STAN. L. REV. 957, 957 (1982).

¹¹⁰ See Craig J. Konnoth, *Revoking Rights*, 66 HASTINGS L.J. 1365, 1378 (2015) (explaining Radin's theory that "property is what permits an individual to project herself into the future or the past [and] . . . populate one's memory with substance").

¹¹¹ See Radin, *supra* note 109 (noting that property rights provide the control needed to achieve proper self-development).

¹¹² President Barack Obama, Remarks by the President at the Precision Medicine Panel Discussion (Feb. 25, 2016), <https://www.whitehouse.gov/the-press-office/2016/02/25/remarks-president>

to individual fitness, health, work, and recreation. For example, data about ovulation and other reproductive information (now often generated through apps) can assist with planning one's family.¹¹³ Information is therefore used to develop a range of life choices in contexts of both work and play. President Obama himself supports an autonomy-based ownership understanding of information: "[O]nce you understand [that the data is] yours, . . . you have agency in the process."¹¹⁴

But apart from autonomy, information about an individual's bodily characteristics, whether or not it is attached to a specific identifier, has deeper ontological significance.¹¹⁵ It helps us conceptualize the person as a person. Information regarding one's DNA helps establish one's connection to one's family, perhaps even to one's ethnic or social group. Key demographic characteristics such as race and gender are constructed in part by—and understood through—medical frames.¹¹⁶ Other health-relevant information, such as sexual behavior and orientation, perform similar functions.¹¹⁷

Further, the "quantified-self" movement promotes data streams as the best form of self-conceptualization and knowledge.¹¹⁸ This movement promotes the use of devices that not only "solve problems related to health," but also produce data about steps walked, heartbeat, calories (consumed and burned),

precision-medicine-panel-discussion [<https://perma.cc/VY3B-CS95>] [hereinafter President Obama, Precision Medicine Panel Remarks].

¹¹³ See Melissa Willets, *The 10 Best Fertility Apps*, FITPREGNANCY.COM, <http://www.fitpregnancy.com/pregnancy/getting-pregnant/best-fertility-apps> [<https://perma.cc/GKQ8-ARAP>] (exploring how various apps can help individuals with family planning).

¹¹⁴ Darius Tahir, *President Talks eHealth*, POLITICO (Oct. 14, 2016, 10:00 AM), <http://www.politico.com/tipsheets/morning-ehealth/2016/10/president-talks-ehealth-fda-guidance-announcement-today-216870#ixzz4Nw5sihqe> [<https://perma.cc/Z7PP-J2F8>].

¹¹⁵ Irma van der Ploeg, *The Body as Data in the Age of Information*, in ROUTLEDGE HANDBOOK OF SURVEILLANCE STUDIES 176, 181 (Kirstie Ball, Kevin D. Haggerty & David Lyon eds., 2012).

¹¹⁶ See generally Rasmus Grønfeldt Winther, *Race and Biology*, in THE ROUTLEDGE COMPANION TO THE PHILOSOPHY OF RACE (Linda Alcoff et al. eds., forthcoming 2017), <https://philpapers.org/archive/WINRAB.pdf> [<https://perma.cc/Y9D6-MZSR>].

¹¹⁷ In a touching story on a prominent blog, one medical professional writes as follows: "Seven years ago, I lost a college-aged patient in a car accident. Placing the final dictation in her chart a week later gave me the opportunity to reflect on our relationship and her assorted illnesses, injuries, and well visits over almost two decades. What a treasure to behold after years of friendship and medical care. Her paper chart was tangible proof of a life well-lived. I recently contacted her mother to inquire if she wanted her daughter's medical chart. She said it was a gift to see her daughter through the eyes of her physician, who was there every step of the way. Medical records are more than metadata on a computer screen; they are a sacred chronicle of our enduring connection with our patients in life, and even in death." Niran Al-Agba, *My Ideal EHR*, HEALTH CARE BLOG (Aug. 23, 2016), <http://thehealthcareblog.com/blog/2016/08/23/my-ideal-ehr/> [<https://perma.cc/Y83T-WMYE>].

¹¹⁸ As anthropologist Dana Greenfield explains, quantified self is a "utopian project[,], where, . . . health behaviors can be changed . . . and self-knowledge, -awareness, and -mindfulness can be achieved." Dana Greenfield, *Deep Data: Notes on the n of 1*, in QUANTIFIED: BIOSENSING TECHNOLOGIES IN EVERYDAY LIFE 123, (Dawn Nafus ed., 2016).

brainwaves, and breathing—to name a few—as a way of knowing oneself.¹¹⁹ Indeed, genetic data is popularly framed as the essence of personhood.¹²⁰ Thus, one surveillance scholar asks, “[W]here exactly is the transition from bodily matter to bodily data? Does it really still make sense to try to make the distinction?”¹²¹ Modern data collection in the form of DNA codes or biomarkers marks a time of transition such that data is not just “representations of the body” but rather “a change on the level of *ontology*.”¹²²

To be sure, as Radin noted of property, not all information will have the same level of import to everyone in all contexts. Property can range from items that are deeply constitutive of personhood to “fungible” items like money that are held merely for instrumental reasons.¹²³ Similarly, data can range from information which embeds and constructs important self-conceptualization to facts that hold little meaning. For example, to an HIV-positive person, a low white cell blood count can have a major impact; to others, it may not. Depending on context, such information can lie on a spectrum.

But pulled together, even trivial data can create a mosaic picture of the individual in which she has deep personhood interests.¹²⁴ Genetic data can help recreate facial structure.¹²⁵ Various health data brought together could predict various personality traits or moods like irritability, depression, generosity, and stress.

Finally, data determines how an individual is seen by others. As Louis Althusser’s famous theory of interpellation explains, a person is constituted, in part, through the characteristics that society and its members attribute to

¹¹⁹ Kashmir Hill, *Adventures in Self-Surveillance, Aka the Quantified Self, Aka Extreme Navel-Gazing*, FORBES (Apr. 7, 2011, 11:34 AM), <http://www.forbes.com/sites/kashmirhill/2011/04/07/adventures-in-self-surveillance-aka-the-quantified-self-aka-extreme-navel-gazing/#1916f7bf7b12> [<https://perma.cc/45E3-C465>]; see also *Counting Every Moment*, ECONOMIST (Mar. 3, 2012), <http://www.economist.com/node/21548493> [<https://perma.cc/QU5G-7Q79>] (exploring the concept of self-tracking and the “belief that gathering and analyzing data about [one’s] everyday activities can help . . . improve [one’s] li[fe]”).

¹²⁰ For example, 23andMe’s test kit reads “Welcome to You,” and an informational webpage is entitled “23 pairs of chromosomes. One unique you.” *23 Pairs of Chromosome, One Unique You*, 23ANDMEBLOG (Aug. 13, 2015), <http://blog.23andme.com/23andme-and-you/23andme-how-to/23-pairs-of-chromosomes-one-unique-you/> [<https://perma.cc/Y9Q2-HA2K>].

¹²¹ van der Ploeg, *supra* note 115, at 180.

¹²² *Id.* at 178-79.

¹²³ See Radin, *supra* note 109, at 960, 966 (describing how some property is fungible to everyone, like money, while some property is fungible to certain individuals, like a car to a dealer or hair to a wigmaker).

¹²⁴ Cf. Orin S. Kerr, *The Mosaic Theory of the Fourth Amendment*, 111 MICH. L. REV. 311, 320 (2012) (suggesting that “five justices are ready to embrace the . . . mosaic approach”).

¹²⁵ See Sara Reardon, *Mugshots Built from DNA Data: Computer Program Crudely Predicts a Facial Structure from Genetic Variations*, NATURE (Mar. 20, 2014), <http://www.nature.com/news/mugshots-built-from-dna-data-1.14899> [<https://perma.cc/9DKF-RFGF>] (describing a new “computer program that can create a . . . 3D model of a face from a DNA sample”).

her.¹²⁶ The collection and analysis of the data categorizes the individual based on her genes, medication, addictions, and disease history as low or high risk, normal or abnormal, healthy or unhealthy. An individual may well feel discomfort at being characterized as abnormal by some faraway anonymous researcher even if that label cannot be ultimately attributed to her.

Even if the data is not used to recategorize the individual *per se*, individuals may object to what is done with their information. Because they retain a bond with their information, they may feel responsible when their information is used for research into procedures that they find offensive such as abortion or blood transfusions.¹²⁷ In a recent qualitative study, individuals also objected to research into “intelligence and race,” which they felt can be used perpetuate past discrimination¹²⁸ and expressed concern that their data may be used by “law enforcement in . . . fishing schemes.”¹²⁹

Thus, what may be at stake for many individuals is an aspect of their personhood. Even if the results of the research cannot be connected back to a particular patient, her sense of dignity as an autonomous individual is at stake.¹³⁰ Therefore, individuals feel they have a claim to their information, whether it is identified or de-identified. As the Agency for Healthcare Research Quality notes, a majority of individuals use the metaphor of ownership to describe their relationship with the data, and a majority object to even de-identified use of the data without consent.¹³¹

C. Expressive Harms

Data collection activity can cause what I call “expressive harms.” The way in which society distributes its resources and treats individuals’ personhood

¹²⁶ Louis Althusser, *On the Reproduction of the Conditions of Production*, in *LENIN AND PHILOSOPHY AND OTHER ESSAYS* 121, 161 (Ben Brewster trans., 1971).

¹²⁷ See Amy R. Applebaum, Note, *When Parental Autonomy Clashes with a Child’s Interest in the Advances of Science: The Case for the Future of Court-Ordered Gene Therapy*, 48 *WAYNE L. REV.* 1543, 1547-48 (2002) (“Many cases involving court-ordered medical treatments of minors have involved the objections of Jehovah’s Witnesses to blood transfusions based on their religion’s mandate forbidding such procedures . . .”).

¹²⁸ Catherine M. Hammack, The Petrie–Flom Ctr. for Health Law Policy, Biotechnology, and Bioethics, 2016 Annual Conference: Big Data, Health Law, and Bioethics: Thought Leader Perspectives on Risks and Protections in Precision Medicine Research (June 16, 2016), <https://www.slideshare.net/petrieflom/catherine-m-hammack-thought-leader-perspectives-on-risks-and-protections-in-precision-medicine-research> [<https://perma.cc/RQ86-QLWX>].

¹²⁹ *Id.*

¹³⁰ Even if information is never improperly released and never results in any concrete harms, information collection is inherently a status-creating activity. Konnoth, *supra* note 7, at 64. As the Supreme Court has recognized several times, information collection has an expressive aspect; it can humiliate or “send a message.” *Id.* at 17, 20, 23.

¹³¹ See WESTAT, *supra* note 108, at 6 (explaining that surveys suggest health care consumers want to “decide what goes into and who can access their medical records”).

symbolically demonstrates respect or disrespect towards them. Welfare recipients are constantly surveilled: social workers intrude into their homes to supervise their family and child-rearing behavior, their food consumption is monitored, and they are subject to random drug tests.¹³² Racial minorities are subject to searches in schools and on the streets, and certain reproductive choices that poor women make are subject to surveillance.¹³³

Placing these individuals under greater degrees of scrutiny than other members of society marks them as less deserving of privacy, and as having less of a right to determine when and how information about them should be released. Those who can control information access, by contrast, are imbued with social status. Subjecting only certain groups of individuals in society to information burdens and exempting others reinforces existing hierarchies in society.¹³⁴

In particular, through the programs I describe, individuals on Medicare and Medicaid are situated within a framework where they are regularly subjected to informational risk and indignity.¹³⁵ By marking low-income individuals in this way, health information surveillance further “otherizes” them from mainstream society and citizenship.

But the kind of surveillance that occurs here imposes a new kind of *exploitative* indignity. Health information is a resource in which individuals have intimate personhood interests, whether the data is identified or de-identified. Moreover, maintaining the material wellbeing necessary for autonomy—employment, health insurance, and engagement with the medical system—also requires informational integrity. Disproportionately sacrificing a group’s privacy, information security, and dignity for the benefit of the population as

¹³² See Bryce Covert, *Drug Testing Welfare Recipients Is a Popular New Policy that Costs States Millions. Here Are the Results*, THINKPROGRESS (Feb. 19, 2016), <https://thinkprogress.org/drug-testing-welfare-recipients-is-a-popular-new-policy-that-cost-states-millions-here-are-the-cf829257ade0#jp2rkxrfn> [<https://perma.cc/262V-A79Z>] (“At least 13 state legislatures have enacted laws to drug test TANF applicants or beneficiaries.”).

¹³³ See Konnoth, *supra* note 7, at 25 (noting that “parents may be informed about the reproductive decisions of their minor children within their custody”).

¹³⁴ See *id.* at 37 (“[I]ntrusions and surveillance are inherently confrontational and hierarchizing.”).

¹³⁵ Intriguingly, a 1990 letter to the *New York Times* suggests that privacy may be a luxury only the wealthy can afford. See Hollis Robbins, Letter to the Editor, *Don’t Let Privacy Be a Luxury for the Wealthy*, N.Y. TIMES (Jan. 4, 1990), <http://www.nytimes.com/1990/01/04/opinion/1-don-t-let-privacy-be-a-luxury-for-the-wealthy-927590.html> [<https://perma.cc/L3KJ-GY8Q>] (describing the prohibitive cost of maintaining a caller ID block); see also Josh Bowers, *Probable Cause, Constitutional Reasonableness, and the Unrecognized Point of A “Pointless Indignity,”* 66 STAN. L. REV. 987, 1038 (2014) (describing searches in economically distressed and historically disadvantaged communities); Dorothy E. Roberts, *Punishing Drug Addicts Who Have Babies: Women of Color, Equality, and the Right of Privacy*, 104 HARV. L. REV. 1419, 1422 (1991) (noting that government intrusion into the lives of pregnant women is particularly harsh for poor women of color); Christopher Slobogin, *The Poverty Exception to the Fourth Amendment*, 55 FLA. L. REV. 391, 392, 406 (2003) (arguing that poor people are afforded fewer privacy and autonomy protections than the wealthy).

a whole, rather than using it only in service of group wellbeing, constitutes a kind of exploitation of one part of the population for the rest. This harms dignity interests—not just because of how data is *collected*, but because of how it is *used*.

This renders the health information collection system different from other kinds of information programs. Welfare, penal, and reproductive surveillance is based on paternalism or stereotyping and grounded in perceptions of difference—of danger, dishonesty, or immaturity.¹³⁶ But the health information collection system is, ironically, based on a perception of biological sameness. Rendering CMS data the basis for population-based medical research makes sense only if one believes that beneficiaries' data is on some level *similar* to that of the rest of the population. This sameness becomes the ground for a kind of exploitation that exempts those who can afford private insurance from the obligations of information collection, while still allowing them access to the benefits. The poor and elderly become proxies in a conscription scheme for the greater good.

This harm is best conceptualized by Julie Cohen's work on the commodification of individuals as data.¹³⁷ Cohen argues that commercial big data ventures seek to "extract . . . marketplace value" from consumers, who are treated as "a repository of raw materials that are there for the taking and that are framed as inputs to particular types of productive activity."¹³⁸ The function of these technologies is anti-individualistic: it is "to subsume individual variation and idiosyncrasy within a probabilistic gradient. Their purpose is to make human behaviors and preferences calculable [and] predictable . . . *in aggregate* . . . [P]artial (or even complete) misalignments at the individual level are irrelevant."¹³⁹

Cohen focuses on the consumer marketplace, but many of her insights can be extended to medical research as well. Despite how laudable the result may be, harvesting medical data shifts its role. Where medical research was once individuated—the means for diagnosis, treatment, and cure of the individual—it becomes alienated and harnessed for the purposes of others. Cohen relies on the famous case *Moore v. Regents of the University of California*, where a patient sued to assert a right to the products made from his cells.¹⁴⁰ Such processes

¹³⁶ Konnoth, *supra* note 7, at 51.

¹³⁷ My solution, of course, only addresses inequality related problems. The broader issues that Cohen identifies may suggest *limiting* the system altogether. As I note in the Introduction, I am not averse to this outcome—as long as whatever outcome is achieved is done so in a way that is equitable.

¹³⁸ Julie E. Cohen, *The Biopolitical Public Domain 2* (Sept. 28, 2015) (unpublished manuscript), <http://ssrn.com/abstract=2666570>.

¹³⁹ *Id.* at 19.

¹⁴⁰ *See id.* at 24 (discussing *Moore v. Regents of the University of California*, 793 P.2d 479 (Cal. 1990), which held that patients do not have a right to share in the profits earned from the research of their cells). Cohen also identifies the harnessing of subordinated populations' (such as

“naturalize[] practices of appropriation” as applied to specific groups such as those on welfare.¹⁴¹

This kind of exploitation undermines autonomy by creating a relationship of domination over poorer and older individuals. Iris Marion Young identifies exploitation as the first of oppression’s faces: exploitation “occurs through a steady process of the transfer of the results of the labor of one social group to benefit another.”¹⁴² Here, it is not the labor of individuals that is being sacrificed, but rather, their privacy, information, medical security, and even their dignity. Rather than deploying resources for their own flourishing and according to their own perceptions of what is good, the poor and elderly are being asked to sacrifice for others.

D. Medical Harms

Finally, for those who remain unmoved by the dignitary harms visited upon individuals because of what they may characterize as uneducated attitudes toward data, these issues raise practical consequences as well. Surveys show that many individuals avoid receiving medical care rather than being subject to health collection mandates. For example, the California Health Care Foundation found that 13-17% of consumers attempt to hide information to protect their privacy.¹⁴³

The avoidance varies by medical condition, and such behavior is usually most prevalent in the early stages of the onset of a condition. Thus, HHS estimated that based on privacy concerns, 586,000 Americans did not seek earlier cancer treatment and that roughly 2,000,000 Americans did not seek treatment for mental illness.¹⁴⁴ Likewise, HHS noted that “there [was] great

welfare recipients) data more generally than discussed here. See Cohen, *supra* note 138, at 11 (“The push to exploit the biopolitical public domain is a contest over a postcolonial terrain, in which global networked elites seek to harness the power of populations worldwide.”).

¹⁴¹ *Id.* at 25.

¹⁴² IRIS MARION YOUNG, *JUSTICE AND THE POLITICS OF DIFFERENCE* 49 (2011 ed. 2011). Bioethics scholarship is of course replete with discussion and definition of exploitation. See, e.g., Ruth Macklin, *Bioethics, Vulnerability, and Protection*, 17 *BIOETHICS* 472, 475 (2003) (discussing the exploitation of vulnerable populations in medical research). But Young better explains the structural kind of exploitation with which I am concerned. An individual’s sense of self, personhood, identity, and capacity to function is, to various degrees, formed, supported, and maintained using social resources, institutions, community norms, symbols, hermeneutics, habits, and tradition. See Simon Caney, *Liberalism and Communitarianism: A Misconceived Debate*, 40 *POL. STUD.* 273, 276-79 (1992) (explaining the “embeddedness thesis,” which states that “it is a sociological fact that a person’s character is shaped . . . by cultural context”).

¹⁴³ CAL. HEALTHCARE FOUND., *NATIONAL CONSUMER HEALTH PRIVACY SURVEY 2005 4* (2005), <http://www.chcf.org/~media/MEDIA%20LIBRARY%20Files/PDF/PDF%20C/PDF%20consumerprivacy2005execsum.pdf> [<https://perma.cc/9FYS-ARVC>].

¹⁴⁴ Standards for Privacy of Individually Identifiable Health Information, 65 Fed. Reg. 82,462, 82,777-79 (Dec. 28, 2000).

uncertainty” as to the incidence of sexually transmitted diseases “due to under-reporting.”¹⁴⁵

These programs, which render the data of the poor and elderly more easily available for secondary research, are in their early stages. To my knowledge, their existence has passed under the radar. But as further research makes the disparate use of this data apparent, along with the concomitant dignity, privacy, and security harms, it might well increase distrust in the public medical system. A single data *breach* of CMS data would be even more devastating.

The effects could last generations among poor and elderly populations. We see analogs elsewhere—the dignitary and material harms faced by racial minorities in the United States, for example, has led to dramatic mistrust of law enforcement. In the medical context, the closest analog might be the Tuskegee Study, which left syphilis untreated in 600 African-American men.¹⁴⁶ For forty years, the men were kept from many forms of treatment for the sake of research without knowing the study existed.¹⁴⁷ Many researchers believe that that study continues to bear significant responsibility for mistrustful attitudes toward medical care and research participation, and for the resultant negative medical outcomes, among African Americans.¹⁴⁸ Although the physical harms it imposes are lesser, the differential treatment of their data can similarly harm the relationship of the poor and elderly with the medical system by decreasing trust.

* * *

The poor and elderly are uniquely vulnerable to the harms I have listed. Financial or medical identity theft can wipe out one’s life savings or lead to delays in receiving medical services. While financial identity theft might take greater *amounts* of money from wealthier individuals in absolute terms, low-income, less-educated, and elderly individuals are less likely to be able to recover from medical or financial identity theft, go without medical care, or pay for medical care out of pocket.¹⁴⁹

¹⁴⁵ *Id.* at 82,778.

¹⁴⁶ Susan M. Reverby, *Introduction: More than a Metaphor: An Overview of the Scholarship of the Study*, in *TUSKEGEE’S TRUTHS: RETHINKING THE TUSKEGEE SYPHILIS STUDY* 1, 1 (Susan M. Reverby ed., 2000).

¹⁴⁷ *Id.*

¹⁴⁸ See generally *id.*

¹⁴⁹ See Jessica Coombs, Note, *Scamming the Elderly: An Increased Susceptibility to Financial Exploitation Within and Outside of the Family*, *ALB. GOV’T L. REV.* 243, 252-53 (2014) (“When the younger generation [suffer financial losses,] they have more of an opportunity to recover, whereas the elderly might never get back on their feet after what has the potential to be such a tremendous loss.”).

Similarly, status and dignity harms may be uniquely difficult for low-income and elderly individuals.¹⁵⁰ Both groups may have status concerns because they may see themselves as dependent. Both groups are subject to relatively greater surveillance and supervision.¹⁵¹ Harnessing their data for the rest of society, rather than for their own good, can therefore be particularly harmful to their dignity.

III. ARGUING FOR EQUITY

So far, I have identified how laws and regulation have facilitated informational inequities. Principles of bioethics help explain the problems with these inequities.¹⁵² This, in turn, has policy ramifications, as research regulations largely look to principles of bioethics for guidance. Thus, agencies from HHS to the Department of Defense, the Department of Education, the Environmental Protection Agency, and the FDA rely on bioethical principles to set research policy in their respective areas.¹⁵³

¹⁵⁰ While some studies consider privacy preferences based on age, these studies cannot control for the fact that younger individuals have yet to experience various adult situations in which privacy is more important. As ACLU lawyer Jay Stanley notes, a teen or young adult may, when asked about the importance of privacy, only think about the question in a world where the intruders are parents, teachers, or other trusted observers. They do not consider employment or insurance discrimination, law enforcement, or other entities. Jay Stanley, *Do Young People Care About Privacy?*, ACLU (Apr. 29, 2013, 10:00 AM), <https://www.aclu.org/blog/do-young-people-care-about-privacy?redirect=blog/technology-and-liberty/do-young-people-care-about-privacy> [<https://perma.cc/YQ7N-G6HW>]. In any case, these studies show that older individuals value privacy more than younger individuals. One might therefore conclude that at least for now, intrusions on the elderly are experienced as more burdensome. To my knowledge, no robust studies have examined privacy preferences based on income.

¹⁵¹ See *supra* text accompanying notes 31–54; see also ELDER LAW: CASES AND MATERIALS 18–26 (Lawrence A. Frolik & Alison McChrystal Barnes eds., 5th ed. 2011) (explaining how the elderly are reliant on the support of others); *id.* at 355–407 (discussing guardianship); *id.* at 409–33 (discussing power of attorney); *id.* at 447–502 (discussing health care surrogate decisionmaking); *id.* at 503–69 (discussing the vulnerability and supervision of elderly with respect to domestic and institutional caregivers). Of course, not all elderly individuals suffer from a loss of privacy, but more elderly individuals are vulnerable in this way than the general population.

¹⁵² No positive legal provision is doctrinally violated by this problem of inequity. Indeed, the federal statutes discussed in Part I pave the way for these inequities.

¹⁵³ See, e.g., *Federal Policy for the Protection of Human Subjects ('Common Rule')*, HHS.GOV, <http://www.hhs.gov/ohrp/regulations-and-policy/regulations/common-rule/> [<https://perma.cc/6CJ8-X9GS>] (“The current U.S. system of protection for human research subjects is heavily influenced by . . . basic ethical principles . . .”); see also *Federal Policy for the Protection of Human Subjects*, 80 Fed. Reg. 53,933, 53,940, 53,942 (proposed Sept. 8, 2015) (discussing the influence of bioethicists in current initiatives and highlighting the “ethical importance of obtaining consent for genomics research”); Response to Comments Received for “The Menlo Report: Ethical Principles Guiding Information and Communication Technology Research” (“The Menlo Report”) for the Department of Homeland Security (DHS), Science and Technology, Cyber Security Division (CSD), Protected Repository for the Defense of Infrastructure Against Cyber Threats (PREDICT) Project, 77 Fed. Reg. 73,669, 73,670 (Dec. 11, 2012) (stating that the Menlo Report, which sets forth ethical guidelines for information security research, is “founded on the Belmont Model, which was originally

Much of this policy is guided by the Belmont Report for the Protection of Human Subjects.¹⁵⁴ The 1979 Report, prepared by the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, has laid out the essential principles of biomedical research for the last four decades. It was prompted in part by the research atrocities perpetrated during World War II and the Tuskegee Study.¹⁵⁵

The Report identifies three principles as fundamental: “respect for persons,” “beneficence,” and “justice.”¹⁵⁶ The principle of respect demands consideration of an individual’s “autonomous” choice.¹⁵⁷ Beneficence requires “secur[ing] the[] well-being” of individuals.¹⁵⁸ Justice considers “fairness in distribution.”¹⁵⁹ Each of these principles is underdefined in the abstract and takes on a shape in particular contexts. The equity problem with secondary research primarily implicates questions of justice.

A generation ago, the Report anticipated some of the issues with which I engage. The Commission observed that in previous “centuries[,] the burdens of serving as research subjects fell largely upon poor ward patients, while the benefits of improved medical care flowed primarily to private patients.”¹⁶⁰ It therefore cautioned that “the selection of research subjects needs to be scrutinized in order to determine whether some classes (e.g., welfare patients . . .) are being systematically selected simply because of their easy availability, [or] their compromised position.”¹⁶¹

For various reasons, however, burden–equity considerations have not traveled into modern health information ethics. First, as a Belmont Report coauthor Albert Jonsen notes, “Justice was the neglected sibling among the principles of bioethics, always acknowledged but seldom given significant tasks or much praise.”¹⁶² In the health information context in particular, the ethical focus is almost exclusively on autonomy concerns, namely, on the

developed for the biomedical research context”); Protections for Subjects in Human Research, 71 Fed. Reg. 6138, 6161 (Feb. 6, 2006) (describing a “principal incentive to conduct research ethically”); Protection of Human Subjects, 62 Fed. Reg. 28,156, 28,156 (proposed May 22, 1997) (noting that the proposed rule is “needed to secure additional protections for children who are involved as subjects of research”).

¹⁵⁴ Belmont Report, *supra* note 12.

¹⁵⁵ King, *supra* note 13, at 137.

¹⁵⁶ See Belmont Report, *supra* note 12, at 23,192 (capitalization omitted).

¹⁵⁷ *Id.* at 23,193.

¹⁵⁸ *Id.* at 23,194.

¹⁵⁹ *Id.* (internal quotation marks omitted).

¹⁶⁰ *Id.*

¹⁶¹ *Id.*

¹⁶² ALBERT R. JONSEN, THE BIRTH OF BIOETHICS 413 (1998).

amount and kind of consent required to collect and use data, and the importance of privacy protections.¹⁶³

Second, justice as it is conceptualized today looks very different than its Belmont avatar. The problem I grapple with raises the question of how to distribute *burdens*. However, as Patricia King—another Report coauthor—notes more recently, “the last thirty years” of bioethics have produced a “dramatic change” where questions of justice are concerned.¹⁶⁴ Ethicists now emphasize “access” to the “*benefits*” of research rather than “protection,” from its “risks and burdens.”¹⁶⁵ If anything, research ethics is concerned with the *lack* of inclusion of minorities and women in randomized control trials, rather than the *overinclusion* of such groups.¹⁶⁶

Third, apart from noting in passing that “social practices such as punishment, taxation and political representation” engage with questions of

¹⁶³ Different groups endorse the weighing of autonomy against health interests. See, e.g., LAWRENCE O. GOSTIN, PUBLIC HEALTH LAW: POWER, DUTY, RESTRAINT 325 (2000) (“Legislation . . . cannot provide absolute privacy protection while still affording reasonable access to data to achieve important public health purposes.”). The outcome of this balancing depends on context. See I. Glenn Cohen et al., *The Legal and Ethical Concerns that Arise from Using Complex Predictive Analytics in Health Care*, 33 HEALTH AFF. 1139, 1144 (2014) (“Predictive analytics falls between two well-accepted models regarding consent. On the one hand, [it] resembles clinical research, in which explicit consent is usually required. On the other hand, it also resembles quality assurance or quality improvement activities, in which consent is not generally required.”); Don E. Detmer, *Your Privacy or Your Health—Will Medical Privacy Legislation Stop Quality Health Care?*, 12 INT’L J. FOR QUALITY HEALTH CARE 1, 2 (2000) (“The issue resembles a teeter-totter with health on one end and privacy on the other. Where one places the fulcrum of law beneath the board is crucial.”). Of course, a key problem is that autonomy-seeking consent approaches do not work very well because patients do not read or understand these policies. See Frank Pasquale, *Grand Bargains for Big Data: The Emerging Law of Health Information*, 72 MD. L. REV. 682, 703 (2013) (noting that most Americans do not understand the payment systems that American hospitals use).

¹⁶⁴ King, *supra* note 13, at 136; see also Madison Powers, *Theories of Justice in the Context of Research* (“A striking consequence of the egalitarian’s focus on goods such as income, wealth, and health care, which persons need to flourish, is that the distribution of burdens and risks can remain largely ignored in the analysis.”), in BEYOND CONSENT: SEEKING JUSTICE IN RESEARCH 147, 150 (Jeffrey P. Kahn et al. eds., 1998); Charles Weijer, *Evolving Ethical Issues in Selection of Subjects for Clinical Research*, 5 CAMBRIDGE Q. HEALTHCARE ETHICS 334, 343 (1996) (noting that “a number of issues, in succession [have been] viewed by ethicists to be the prime concern in the selection of subjects for clinical research”).

¹⁶⁵ King, *supra* note 13, at 136 (emphasis added).

¹⁶⁶ See *id.*; see also Carol Levine, *Changing Views of Justice After Belmont: AIDS and the Inclusion of “Vulnerable” Subjects* (“[R]esearchers . . . should not offer potentially beneficial research only to some persons who are in their favor . . .” (internal quotation marks omitted)), in THE ETHICS OF RESEARCH INVOLVING HUMAN SUBJECTS: FACING THE 21ST CENTURY 105, 107 (Harold Y. Vanderpool ed., 1996). Many argue that the shift resulted due to the legacy of HIV, where certain groups were excluded from participation in HIV clinical trials in which the best medication was available. See, e.g., Robert J. Levine, *The Impact of HIV Infection on Society’s Perception of Clinical Trials*, 4 KENNEDY INST. ETHICS J. 93, 95 (1994) (“AIDS activists . . . correctly pointed out that enrollment in [the azidothymidine drug] clinical trial was the only way for HIV-infected persons to get even a 50 percent chance of receiving the only therapy that offered any hope of delaying death or the onset of opportunistic infections.”).

burden–equity, the Report does not provide a burden–equity analysis.¹⁶⁷ Since the literature has not had occasion to engage with the question since then, a robust explanation is required.

A. Health Information Norms

In a 2016 discussion about precision medicine and the learning health system, President Obama explained that we are building a shared communal resource: “all of us . . . could have electronic medical records that . . . we pool together . . . that researchers, practitioners, [and] scientists can share.”¹⁶⁸ This would require engagement from all segments of the population: “from the private sector, from the public sector, from the non-for-profit sector, from the medical community, [and] from researchers.”¹⁶⁹

This shared purpose draws on intuitions of reciprocity. The President was quick to emphasize that, as an initial matter, the data belonged to the patient: “I would like to think that if somebody does a test on me or my genes, that that’s mine.”¹⁷⁰ But by providing the data that they owned, individuals would be repaid in medical benefits, and their contributions would allow us to “individualiz[e] treatments for a particular patient.”¹⁷¹ The system will “give all of us access to the personalized information we need to keep ourselves and our families healthier.”¹⁷² Thus, even as all individuals contribute to the system, all individuals draw from it in various ways

The fairness and reciprocity that undergirds the President’s description has broad intuitive appeal. Both popular and moral accounts deploy this kind of reasoning in other contexts to justify pooling of resources. For example, scholars and politicians have used solidarity- and reciprocity-based approaches to justify tax policy.¹⁷³

¹⁶⁷ Belmont Report, *supra* note 12, at 23,194.

¹⁶⁸ President Obama, Precision Medicine Panel Remarks, *supra* note 112.

¹⁶⁹ *Id.* A former FDA Commissioner, Robert Califf, even suggested that it is a “patriotic” duty to provide information. Darius Tahir, *Administration with Big PMI Moves*, POLITICO (July 7, 2016, 10:00 AM), <http://www.politico.com/tipsheets/morning-ehealth/2016/07/administration-with-big-pmi-moves-congressional-roundup-cms-with-meaningful-use-tweaks-215208> [<https://perma.cc/MF> G5-G6ER] (internal quotation marks omitted).

¹⁷⁰ President Obama, Precision Medicine Panel Remarks, *supra* note 112.

¹⁷¹ *Id.*

¹⁷² President Barack Obama, Remarks by the President in State of the Union Address (Jan. 20, 2015), transcript available at <https://www.whitehouse.gov/the-press-office/2015/01/20/remarks-president-state-union-address-january-20-2015> [<https://perma.cc/QDB2-WMPM>].

¹⁷³ See Joseph M. Dodge, *Theories of Tax Justice: Ruminations on the Benefit, Partnership, and Ability-to-Pay Principles*, 58 TAX L. REV. 399, 406 (2005) (explaining the new benefit principle of taxation, which “purports to be a norm of tax fairness that . . . postulat[es] that the measure of a person’s benefit from government is none other than his or her financial . . . well-being”); *American Community Survey (ACS): Is the American Community Survey Mandatory?*, U.S. CENSUS BUREAU, <https://www.census.gov/programs-surveys/acs/about/survey-is-mandatory.html> [<https://perma.cc/X>].

Everyday moral reasoning assesses a particular activity based on the norms of the context or institution in which it is situated—the particular law school, family, or community. We decide if that is “the way we do things here.” Thus, here, we consider only the processes generally attributed to the health information system. We do not consider processes, resources, or people’s social situations outside the system. Adopting this internal point of view allows us to provide a *prima facie* assessment of the justice of the policies within the health information system without grappling with the complications of a broader perspective that accounts for other social institutions.

We know that contemporary health information policies facilitate a world in which the poor and elderly, as a group, will provide most of the health information. We also know they may enjoy unique benefits from the health information system. Health information research, as explained above, can roughly be divided into health care delivery research, carried out using claims data, and clinical research, carried out using EHR data.¹⁷⁴

Medicare and Medicaid are unique public health insurance programs with health care delivery that can be *sui generis* due to their size and structure. The claims data from these programs might reflect this uniqueness, and research using that data could conceivably provide primary benefits for the programs and their beneficiaries.

But *clinical* research that CMS policies will eventually produce will accrue to everyone’s benefit.¹⁷⁵ While there may be some conditions that are more common among the poor and elderly, there is no major condition that is unique to those groups of individuals. Adverse health events impact everyone. Yet, many

DJ8-W6MZ] [hereinafter ACS Survey] (emphasizing social unity to describe why the American Community Survey is mandatory); see also Lawrence Zelenak, *Taxing Endowment*, 55 DUKE L.J. 1145, 1153 (2006). In fact, through regressive consumption taxes and graduated income taxes, President Lincoln sought to send the message that “all Northern citizens were supporting the war equally,” and Republican lawmakers “employ[ed] the rhetoric of patriotism and shared sacrifice to frame the paying of new robust taxes as a form of loyalty to the Union.” Ajay K. Mehrotra, *The Price of Conflict: War, Taxes, and the Politics of Fiscal Citizenship*, 108 MICH. L. REV. 1053, 1060–61 (2010). Similarly, as historian Carol Jones argues, although dramatic expansion of the tax base during World War II hurt the less well-off, it sent a solidarity-driven message: all citizens shared the cost of war. See Carolyn C. Jones, *Mass-Based Income Taxation: Creating a Taxpaying Culture, 1940–1952* (explaining that tax-focused “propaganda messages throughout the war used a theme of comparative sacrifice”), in *FUNDING THE MODERN AMERICAN STATE, 1941–1995*, at 107, 114 (W. Elliot Brownlee ed., 1996); see also Reuven S. Avi-Yonah, *The Three Goals of Taxation*, 60 TAX L. REV. 1, 7 (2006) (noting that many governments used World War II “to greatly expand their income tax collections”).

¹⁷⁴ See *supra* text accompanying notes 26–28.

¹⁷⁵ The harms to which I point are not immediate, of course. The system has not yet completed the process of making data available. Benefits are further down the line. See Mike Orcutt, *The White House Is Pushing Precision Medicine, but It Won’t Happen for Years*, MIT TECH. REV. (July 18, 2016), <https://www.technologyreview.com/s/601883/the-white-house-is-pushing-precision-medicine-but-it-wont-happen-for-years/> [<https://perma.cc/TMA4-PM57>] (“[W]e are many years from realizing [the] ‘new era of medicine’ the [P]resident described in his 2015 State of the Union address . . .”).

individuals who are richer or younger will gain access to benefits without their data ever being used in clinical research. And many have already raised concerns that fewer benefits will accrue to the worse off as time goes on.¹⁷⁶

And even when it comes to claims data, there is evidence to show that commercial entities are seeking to use the data to improve the lot of the richer from whom they can expect better remuneration. Take, for example, the research program I describe above arising from the ACA and subsequent 2015 legislation.¹⁷⁷ The ACA program originally forbade using data for purposes other than CMS quality improvement (which excluded, for example, generalizable secondary research that could benefit non-CMS beneficiaries as well).¹⁷⁸ Notwithstanding the statutory ban, research entities sought greater freedom after the proposed rules were issued so that they could market a greater range of private analyses that did not pertain to CMS quality improvement.¹⁷⁹ CMS, of course, pointed out that the statute prohibited such broader uses.¹⁸⁰

But the 2015 legislation changed the ACA standard to allow qualified entities to perform any analysis, including secondary research for certain data suppliers—such as providers, employers, and insurers.¹⁸¹ CMS noted that it expected that the number of entities applying to use the data to increase due to the more permissive 2015 regime.¹⁸² The policies therefore facilitate an outcome that does not conform to basic intuitions of reciprocity and fair play, by using the data of only CMS beneficiaries for everyone’s potential benefit.

To be sure, the system may not capture everyone’s data. One-and-a-half percent of individuals will never interact with the medical system¹⁸³ and therefore may never gain from these data analyses. But the vast majority of society should give back through health information contributions to the system from which they draw. The only way to maintain the shared stake that demands reciprocity is by requiring equitable contributions from all individuals.

¹⁷⁶ See, e.g., Barbara Feder Ostrov, *The Challenge of Taking Health Apps Beyond The Well-Heeled*, NPR (June 23, 2016, 8:58 AM), <http://www.npr.org/sections/health-shots/2016/06/23/483098999/the-challenge-of-taking-health-apps-beyond-the-well-heeled> [<https://perma.cc/W25W-3Y6F>] (discussing the importance of “tailor[ing] digital health technologies to lower-income people not only to be fair, but because they’re more likely to have chronic illnesses . . . that are expensive to treat”).

¹⁷⁷ See *supra* note 47 and accompanying text.

¹⁷⁸ Medicare Program; Availability of Medicare Data for Performance Measurement, 76 Fed. Reg. 76,542, 76,542 (Dec. 7, 2011).

¹⁷⁹ *Id.*

¹⁸⁰ *Id.*

¹⁸¹ See Medicare Program: Expanding Uses of Medicare Data by Qualified Entities, 81 Fed. Reg. 5397, 5415 (Feb. 2, 2016) (to be codified at 42 C.F.R. pt. 401) (increasing access for nonpublic analyses of data); see also *id.* at 5400 (discussing the details of how to identify providers and suppliers of these now expanded analyses).

¹⁸² *Id.* at 5398.

¹⁸³ See CDC STATISTICS, *supra* note 75.

B. General Social Norms

Although everyday moral reasoning starts out with accounts internal to specific contexts, those contexts are embedded within larger social structures. Reflection often involves abstracting up and testing the robustness of our intuitions against sets of moral norms that apply more generally. When assessing a certain behavior, it is not enough to decide whether “it is the way we do things here”; we also assess how the behavior comports with broader social values about justice and fairness.

An example of this is the taxation system. In the abstract, one could argue that the central goal of taxation is merely to provide for the upkeep of government. Norms of equity and reciprocity may seem to counsel a flat tax system with everyone contributing equally. But our existing system considers a range of values that borrow from various other social norms such as justice, equity, family, and mutual obligation.¹⁸⁴ We consider the position of individuals in society more generally—their income, health, age, and family size—and we modulate contributions as well as payouts.¹⁸⁵ Those with higher incomes receive less and pay more.

184 See C. Eugene Steuerle, *And Equal (Tax) Justice for All?* (discussing seven adjustments made to ensure equity in tax contributions including “income in-kind, potential income or consumption, need, transfers paid and received prices, household size, and measurement period”), in *TAX JUSTICE: THE ONGOING DEBATE* 253, 273-78 (Joseph J. Thorndike & Dennis J. Ventry Jr. eds., 2002). Even if one just used income, what income metric should be used? See generally Alan J. Auerbach & Kevin Hassett, *A New Measure of Horizontal Equity*, 92 AM. ECON. ASS’N 1116 (2002), <http://www.nber.org/papers/w7035.pdf> [<https://perma.cc/LS2K-U4GK>] (discussing available models by which to measure income); Daniel Markovits, *Luck Egalitarianism and Political Solidarity*, 9 THEORETICAL INQUIRIES L. 271 (2007) (defending the theory of luck egalitarianism, which prioritizes personal responsibility in the context of redistributive justice). For example, even with as fundamental a measure as income, some argue that *potential* income is a better measure than *actual* income. See Steuerle, *supra*, at 274. The idea is that individuals should be compared based on the potential income they may earn for fairness reasons. An individual who is able to earn more income and contribute more to society—but chooses not to—should pay more than an individual whose maximum potential is lesser, even if the latter actually earns more than the former. Second, those who take such a position may argue that taxing actual income may disincentivize individuals from maximizing income. But others respond that we cannot ethically force individuals to earn their optimum income. Taxing the potential income of individuals who earn suboptimally is unethical under the redistributive and benefit principles, and also impractical—a fool’s errand that seeks to collect money that isn’t there. Finally, measuring actual income is feasible; measuring potential income is not. *Id.* at 275. Accordingly, we rely on actual rather than potential income to assess taxes.

Similar to the income context, some may argue that potential—rather than actual health—is the better criterion to use when calculating welfare. We should not treat an individual who does not take care of her health with the same care as someone who suffers through no choice of her own. But, as in the income context, we cannot always force individuals to engage in behaviors that are optimal for their health. And it is then arguably unethical (though perhaps not impractical) to increase burdens on those with poor health. Finally, measuring actual health is feasible. Measuring potential health under current technology is not.

185 Steuerle, *supra* note 184, at 274.

We do not always consider extra-contextual values. Some settings, such as the military, are insulated from society as a whole such that broader norms and standards of society do not always apply. But the health information system is not such a setting; it is embedded within a larger social context from which it draws and to which it contributes. The system relies on technology developed in other contexts, as well as the taxes and premiums that support the medical apparatus as a whole. Individuals who participate in those areas indirectly support the health information system. In turn, as President Obama noted, the system contributes back to society: it “promises to reduce costs, provide much better care, [and] make our entire health care system much more effective.”¹⁸⁶ This allows society to focus its resources in other areas from which we benefit.

Even those who never interact with the system may therefore benefit from it indirectly.¹⁸⁷ A law professor who never needs the doctor still depends on the productivity of her colleagues and law review editors.¹⁸⁸ The time she can devote to her work depends in part on whether society can help address any poor health experienced by those in her charge.¹⁸⁹

Further, broader social contexts give us referents to *value* the contributions that the system makes. How to value a cure, for example, depends largely upon the real world effects of a disease on human functioning—which may differ from human to human—and its prevalence. Some societies might penalize certain kinds of bad health or devote resources to projects other than improving the health of the unwell; others attempt to diminish hardship, invest in health care, value jobs that require one to be able-bodied, and so on.¹⁹⁰

¹⁸⁶ President Obama, Precision Medicine Panel Remarks, *supra* note 112.

¹⁸⁷ This responds to an objection by Iain Brassington against obligations for research contributions that has remained unaddressed in the bioethics literature. See Iain Brassington, *John Harris' Argument for a Duty to Research*, 21 *BIOETHICS* 160, 162 (2007) (“[T]here is no problem with free riding when the benefits of research that I enjoy are paid for by some sort of insurance.”).

¹⁸⁸ See ALLEN JENKINS & RON KONECNY, *NEBRASKA MEDICAID EXPANSION* 33-35 (2015), http://www.nebraskahospitals.org/file_download/inline/9eb5a4d7-8725-4385-959f-of404c895128 [<https://perma.cc/UZU5-TWMV>] (“Programs that improve the health and vitality of workers improve the quality of labor.”).

¹⁸⁹ As Bill Sage notes, “One person’s malady can harm families, workplaces, clubs, churches, and sometimes entire communities.” William M. Sage, *Solidarity: Unfashionable, but Still American*, in *THE HASTINGS CTS., CONNECTING AMERICAN VALUES WITH HEALTH REFORM* 10, 10 (Mary Crowley ed., 2009).

¹⁹⁰ See NORMAN DANIELS, *JUST HEALTH: MEETING HEALTH NEEDS FAIRLY* 57 (2008) (arguing that Rawls’s conception of primary social goods can be broadened to include “health-care institutions among the basic institutions involved in providing for fair equality of opportunity”); see also Chai R. Feldblum, *Rectifying the Tilt: Equality Lessons from Religion, Disability, Sexual Orientation, and Transgender*, 54 *ME. L. REV.* 159, 184-85 (2002) (arguing that “we should establish a fund that we all contribute to equally [to] . . . cover[] the costs of [accommodating] any characteristic for which the law has created a reasonable accommodation mandate”).

As the health information system is embedded within a set of larger social structures, we should consider theories of political obligation that apply to society more generally. These theories tell us how an individual's place in society affects and constructs the kinds of claims she can make on society as a whole, and the duties she owes. Those claims and obligations, in turn, permeate all social contexts and affect assessments of fairness internal to them, including those in the health information system.

Although this is not the place to advance a fully outlined theory of political obligation, many such existing, mainstream theories support the conclusion that data collection and its use for secondary research should be carried out equitably.¹⁹¹ Some would argue that those who are richer or younger gain a lot from society and therefore should pay society back and that a poor and elderly individual owes less. Assisting more vulnerable groups also accords with approaches that require that all individuals have certain basic capabilities and autonomy. Theories of general utilitarianism may also require redistribution toward the more vulnerable to improve social welfare as a whole because the marginal improvements to the more vulnerable exceed the losses to the better off. I do not pick among these theories, beyond noting that their conclusion—that we should redistribute resources to assist the worse off in many cases—undergirds many social and economic policies.

Indeed, some would claim that to achieve equity, we should redistribute risk from the more vulnerable to the less vulnerable. Such a claim relies on two premises. Many believe that society currently does not do enough to help

¹⁹¹ Such theories are usually far more complex than the simple, intuitive, reciprocity claims that I deploy—reciprocity by itself might prove indeterminate without some theory concerning baseline rights and duties. Many political theorists, of course, share this intuition. Some more famous approaches include Rawls's difference principle, see J.E.J. Altham, *Rawls's Difference Principle*, 48 *PHILOSOPHY* 75, 75 (1973) ("The difference principle states that the long-run expectations of the least advantaged social group should be maximized."), and Amartya Sen's and Martha Nussbaum's capabilities approaches. See MARTHA C. NUSSBAUM, *WOMEN AND HUMAN DEVELOPMENT: THE CAPABILITIES APPROACH* 5 (2000) ("[T]he best approach to this idea of a basic social minimum is provided by an approach that focuses on *human capabilities*, that is, what people are actually able to do and to be . . ."); AMARTYA SEN, *THE IDEA OF JUSTICE* 231 (2009) (outlining the capability approach, which "in contrast with the utility-based or resource-based lines of thinking, [notes that] individual advantage is judged . . . by a person's capability to do things he or she has reason to value"). Moreover, communitarian theories might impose duties to assist other members of the community. See, e.g., Per Bauhn, *The Extension and Limits of the Duty to Rescue*, 3 *PUB. REASON*, June 2011, at 39, 46 ("[W]e have a duty of necessity to rescue someone whose basic well-being is endangered . . ."). A benefit theory might suggest that the rich gain more from living in society so they owe more back to it. See YAEL TAMIR, *LIBERAL NATIONALISM* 132 (1993) (discussing the theory of political obligations which views them as "way[s] of paying the debts one owes the state, following the enjoyment of benefits and services one has received from it"). See generally JOHN HORTON, *POLITICAL OBLIGATION* (2010).

vulnerable individuals—at least low-income individuals.¹⁹² This claim is both descriptive and normative. As a normative matter, the claim first assumes that there is a particular level of assistance and basic goods that is owed to all individuals. As a descriptive matter, the claim next asserts that this society does not meet that obligation. Society therefore has an obligation to alleviate the burdens on the poor and elderly through available means, including through the health information system.

Each of these intuitions is subject to challenge. First, my normative claim that we must help those who are vulnerable may be questioned. Countervailing theories assert that those who are worse off in society are owed nothing.¹⁹³ Such theorists may indeed argue that since they are owed nothing in the first place, those who receive Medicare and Medicaid are getting a windfall, and—if anything—should be repaying society. Those with this set of moral priors will not find my claims here persuasive, but making a case against their foundational claims is beyond the scope of this Article.

Others might accept the normative claim that we should help the vulnerable, but might reject the descriptive claim, believing that we help the vulnerable too much or just enough. Those who believe we are helping the vulnerable *too much* would likely approve of the current system to some extent. Like those who reject my normative premise, they may see health information inequity as an appropriate way of extracting repayment. Those who believe that we already help the vulnerable just enough would not want the health information system to disturb the status quo. Vulnerable individuals do not need to *repay* anything, but they also do not deserve any additional solicitude. For this group then, the additional burdens (and benefits) that come from the health information system should be distributed equally. But even for this group, the existing system, which distributes burdens unequally, should be adjusted.

C. Correcting Inequity

Inequity can be remedied in three ways. First, we may “level up,” eliminating the new data collection programs and alleviating the burden on everyone. Second, we may compensate by providing the harmed group with an offsetting benefit. Third, we may “level down” by imposing the burden on everyone.

The Introduction notes that the first solution is not off the table. Many believe that enhanced data collection is morally problematic or socially inefficient and that it should be stopped. As a normative matter, I believe that

¹⁹² For examples of scholars that rely on the frameworks that I have discussed to make such an argument, see DANIELS, *supra* note 190, at 15-16, and WIEBKE KUKLYS, AMARTYA SEN'S CAPABILITY APPROACH: THEORETICAL INSIGHTS AND EMPIRICAL APPLICATIONS *passim* (2005).

¹⁹³ See, e.g., ROBERT NOZICK, ANARCHY, STATE, AND UTOPIA 193-94 (1974).

there is no perfectly clear answer here—the relative weight of privacy versus health will depend on an individual’s history, circumstances, and cultural background. Procedural mechanisms such as the regular democratic process, which helps address clashing and incommensurate values and priorities among individuals, will determine the direction in which we must go.

As a practical matter, that process—for the time being and with bipartisan support—has put in place a system that will result in enhanced data collection and use.¹⁹⁴ It is highly unlikely that we will reverse course. But it is entirely possible that including richer individuals within the information dragnet will create resistance to the program, thereby changing the direction of information policy and reducing collection for all. Although I refer to this outcome as “leveling up,” it will actually affect a redistribution of wellbeing. Those who are not on public benefit programs would no longer be able to harness the advantages of research with the data of public beneficiaries. While those *on* the programs would also lose out on the benefits, they would also no longer be burdened and would therefore be relatively better off. This would be a sufficient solution to the equity problems I raise here.

Second, one may offer offsets to individuals who suffer burdens. Indeed, some may say that individuals on public programs already receive offsets in the form of free or reduced-cost medical care. Further, we may also reasonably conclude that the analyses that come from the data will be most relevant to—and most beneficial for—the groups who supply the data.¹⁹⁵

However, neither argument carries weight. The first argument adopts an inappropriate baseline for determining burdens and benefits. The previous Section assumes a general theory of social obligation, under which we owe duties to those who are poor, elderly, or otherwise vulnerable:¹⁹⁶ the beneficiaries of public programs therefore receive no more than is their due under benefit programs. It is thus inaccurate to portray these as offsets.¹⁹⁷

Next, while public benefit recipients do benefit from the outcomes of these analyses, I have explained at length why they are not the only ones to benefit.¹⁹⁸ For an offsetting benefit to create equity, it should only be given to those who suffer the harm. Distributing it across society, even to some degree, would render everyone better off, but still unequal.

¹⁹⁴ See Brady Letter, *supra* note 56, at 2 (“Access to protected [Medicare] data serves as a resource to look into the Medicare program’s costs, services, and trends.”).

¹⁹⁵ See *supra* text accompanying notes 174–75.

¹⁹⁶ See *supra* Section III.B.

¹⁹⁷ It is also incorrect to argue that the worse off are, by incurring risk for the good of others, “paying back” into the system. It is only just to obligate individuals to “pay back” if they owe something in the first place. But since the worse off obtain benefits because of independent moral considerations, they do not owe society anything—it is unjust to force them to repay what they do not owe.

¹⁹⁸ See *supra* text accompanying notes 175–76.

Even if one concedes that all of society unfairly benefits from the burdens on Medicare and Medicaid recipients, to restore fairness, one might seek ways to ensure only that those who give data ultimately get access to the results of the analyses. But how could we implement this? For example, imagine that a doctor learns information about treating diabetes in children of a certain age. It is both impractical and unethical for her to apply this information when treating a child on Medicaid but ignore it when treating a diabetic child who does not receive Medicaid.

But perhaps we could provide Medicaid and Medicare recipients other kinds of compensation for taking their information. In principle, that approach would bring us closer to equity. One could imagine a world in which, in return for a Medicare or Medicaid recipient's health data, we provide more medical benefits, food stamps, or other benefits, according to the recipient's need. To prevent selection biases, we could not make this a consent-based program, as some groups are more willing to provide their data—whether or not in return for benefits—than others.¹⁹⁹

While this approach would create a more equitable world, I remain troubled by its symbolic, expressive effects.²⁰⁰ Why, instead of spreading the data collection burdens across the population at minimal cost (as I propose below),²⁰¹ would we choose to focus them on a vulnerable group of individuals and take their genetic, mental, or sexual history data? To what extent would such disparate policies reinforce differences between social groups? Even if the in-kind benefits outweigh any burdens at the individual level, one could reasonably believe that there may be harms at group level. There may be long-term harms to the body politic even if an individual enjoys short term gains.²⁰²

¹⁹⁹ See *infra* note 205 and accompanying text.

²⁰⁰ See *supra* Section II.C.

²⁰¹ See *infra* Section IV.B.

²⁰² I limit my hypothetical to one where we provide recompense in kind rather than in money because money is symbolically treacherous. Paying for something as intimately connected to an individual as health information risks increasing commodification of an important resource. *Cf., e.g.,* RICHARD M. TITMUS, *THE GIFT RELATIONSHIP: FROM HUMAN BLOOD TO SOCIAL POLICY* 158-72 (Vintage Books ed., 1972) (arguing against the commercialization of blood donation). Both the research ethics literature and federal regulations treat monetary payments as posing autonomy concerns in many instances. See Christine Grady, *Payment of Clinical Research Subjects*, 115 J. CLINICAL INVESTIGATION 1681, 1683-84 (2005) ("Some worry that individuals with limited opportunities for earning money may be most susceptible to impaired judgement when faced with an offer of money . . ."); see also ROBIN LEVIN PENSLAR & JOAN P. PORTER, OFFICE FOR HUMAN RESEARCH PROTS., INSTITUTIONAL REVIEW BOARD GUIDEBOOK ch. III, § G (2008). However, those claims are highly contested in the literature. Others have suggested that aversion to such commercial transactions is badly conceived. See, e.g., Stephen Wilkinson, *Commodification Arguments for the Legal Prohibition of Organ Sale*, 8 HEALTH CARE ANALYSIS 189, 197-98 (2000) (arguing that commodification concerns over payments in exchange for organs are misplaced).

The third approach to achieving equity is to use information from everyone for analyses. This “leveling down” approach would redistribute burdens toward those not on CMS programs. Assume, for example, that a particular research project requires one million records. The current framework would enable the project to draw all the records from those on Medicare and Medicaid, subjecting them to material and dignitary risks and harms. The solution I adopt in the next Part would require the project to draw some of these records from those not on CMS programs, easing the burden on CMS beneficiaries as a group.

D. *A Utilitarian Alternative*

Some may have wondered why I have not simply adopted a utilitarian approach to reach this solution, rather than relying on intuitions that not all might share. A utilitarian approach balances the burden to the individual (often measured both objectively and subjectively) against the benefit to health across society in general.²⁰³ A purely utilitarian approach will assert that contributions that optimize health across society at the least possible cost to the individual are just—though more complex variations are possible. Nonetheless, as we shall see, most varieties of a utilitarian principle prove inconclusive for my purpose, so I do not explore them in depth.²⁰⁴

A utilitarian approach would optimize the process and maximize research benefits while minimizing burdens on individuals. Maximizing benefits might require us to collect data across social groups to have a well-functioning health information system. Although data from public benefit programs may be obtained cheaply, the data is also heavily biased toward elderly and low-income individuals, and such bias can taint research.²⁰⁵

For example, research shows that some social determinants may predict health outcomes better than genetic factors.²⁰⁶ Social conditions can also alter

²⁰³ Risks must be “reasonable in relation to anticipated benefits, if any, to subjects, and the importance of the knowledge that may reasonably be expected to result.” 45 C.F.R. § 46.111(a)(2) (2005); *see also* Standards for Privacy of Individually Identifiable Health Information, 65 Fed. Reg. 82,462, 82,464 (Dec. 28, 2000) (noting that HIPAA balances the “needs of the individual with the needs of society”).

²⁰⁴ *See infra* text accompanying note 208.

²⁰⁵ Much of the literature on secondary health research has focused on the problem of bias—individuals in minority and various other groups are less likely to provide their information, which generates biases in statistical studies. This prevents scientists from drawing robust conclusions regarding disease profiles and the effects of various behavioral, environmental, and other characteristics on health. *See, e.g.,* NASS ET AL., *supra* note 4, at 209–14 (noting that “many small health care entities . . . serving disadvantaged populations are not participating in research[,] . . . [which] results in the underrepresentation of minority populations in many research studies”).

²⁰⁶ *See* Dorothy E. Roberts, *Debating the Cause of Health Disparities: Implications for Bioethics and Racial Equality*, 21 CAMBRIDGE Q. HEALTHCARE ETHICS 332, 333 (2012) (“It has been firmly

genetic expression, confounding the research pool.²⁰⁷ Accordingly, relying on data that is biased towards certain social groups can have problematic effects. Thus, the current system of collection might yield suboptimal social goods—biased data—relative to its autonomy harms. To prevent this, we might collect data from all individuals—or at least a representative sample. For example, the United States Census Bureau collects data from only a small sample of Americans each year.²⁰⁸

Yet, I do not delve deeply into a utilitarian approach because of a key problem—we simply do not have enough information to properly weigh costs and benefits. Although more data will be valuable, collecting data from non-CMS populations could be expensive. If the government is the collecting entity, it must regulate private entities' data quality, compel collection or sampling, and then combine the data for research. And there will be coordination costs if private data collectives run such programs. Incurring these costs without knowing the exact value gained from particular kinds of research is problematic. While social determinants of health are important, so are genetic determinants. It may therefore be cheaper to rely solely on CMS data and focus on genetic or other kinds of research in which biases in the dataset do not matter as much. Thus, although a utilitarian approach may well favor equity, given our current knowledge, I cannot predict how it will eventually cut.²⁰⁹

established that the best predictor of health is an individual's position in the social hierarchy. Hundreds of studies tracking the health of people along the social ladder show that health gradually worsens as status declines.”); see also Harry J. Heiman & Samantha Artiga, *Beyond Health Care: The Role of Social Determinants in Promoting Health and Health Equity*, HENRY J. KAISER FAM. FOUND. (Nov. 4, 2015), <http://kff.org/disparities-policy/issue-brief/beyond-health-care-the-role-of-social-determinants-in-promoting-health-and-health-equity/> [<https://perma.cc/K5KC-EQDC>] (discussing how the ACA can help expand access to health coverage for underserved populations); *NCHHSTP Social Determinants of Health: Frequently Asked Questions*, CDC, <https://www.cdc.gov/nchhstp/socialdeterminants/faq.html> [<https://perma.cc/HUB5-YQ6X>] (last updated Mar. 21, 2014) (discussing the numerous social determinants of health, such as education and income).

²⁰⁷ See Alison Gopnik, *Poverty's Vicious Cycle Can Affect Our Genes*, WALL ST. J. (Sept. 24, 2014, 10:10 AM), <http://www.wsj.com/articles/genes-play-a-role-in-poverty-1411567833> [<https://perma.cc/UU7Y-F75W>] (“Twenty percent of American children grow up in poverty, and this number has been rising, not falling. Nearly a million are maltreated. The new studies show that this damages children, and perhaps even their children's children, at the most fundamental biological level.”).

²⁰⁸ See U.S. CENSUS BUREAU, AMERICAN COMMUNITY SURVEY DESIGN AND METHODOLOGY (JAN. 2014): CHAPTER 4: SAMPLE DESIGN AND SELECTION 1 (2014), https://www2.census.gov/programs-surveys/acs/methodology/design_and_methodology/acs_design_methodology_ch04_2014.pdf [<https://perma.cc/F4ZD-WTDN>] (“Each year . . . , we selected approximately 2.9 million [housing unit] addresses in the [United States].”).

²⁰⁹ Further, even if we had all the relevant facts, the cost-benefit analysis faces problems of valuation and time horizon. What, for example, is the appropriate timeframe? In assessing the value of longitudinal data, should we consider the potential benefits that will accrue ten years from now? One-hundred years from now? How exactly should we value individual autonomy in the balance? Will those values remain the same over time?

IV. USING LAW TO ACHIEVE EQUITY

Law provides a discrete set of policy levers to achieve equity. First, adjusting Medicare and Medicaid, all-payer claims databases, and HIPAA regulations and policy can alleviate collection inequity. Next, additional regulations should be introduced to alter whose data is used in research and analysis. Lastly, all of the recommendations I offer require a restructuring of health information coordination tasks within HHS. Unless otherwise stated, I take it for granted that none of the solutions I offer present any *constitutional* challenges under existing law.²¹⁰

A. Collection

1. EHR Data

CMS can require *EHR* data reporting from patients with private insurance. The Health Information Technology for Economic and Clinical Health (HITECH) Act of 2009 gives CMS the authority to decrease fees for “covered . . . services” if a provider is not a “meaningful EHR user.”²¹¹ The Secretary has broad discretion to determine whether the use of data is “meaningful.” Meaningful use includes “using certified EHR technology in a meaningful manner,” and reporting on measures using EHR “in a form and manner specified by the Secretary, on such . . . measures as selected by the Secretary.”²¹² The Secretary may “requir[e] more stringent measures of meaningful use” as time progresses.²¹³

The Secretary also has broad discretion to define what makes a provider a meaningful user of health information. While she can adjust payments for services provided under Medicare and Medicaid, the reporting does not need

²¹⁰ In the health context, the Supreme Court has shown little appetite for questioning policy decisions regarding health information collection under the Constitution. In *Whalen v. Roe*, the closest on-point case, the Court rejected just such a challenge. See 429 U.S. 589, 598 (1977) (upholding the constitutionality of a state statute that required records of prescriptions for certain dangerous drugs be sent to the Department of Health as “a reasonable exercise of [the state’s] broad police power”); see also *Citizens for Health v. Leavitt*, 428 F.3d 167 (3d Cir. 2005) (rejecting a due process challenge to the HIPAA routine data use exception).

²¹¹ 42 U.S.C. § 1395w-4(a)(7)(i) (2012); see also *id.* § 1395w-4(o)(1)(A) (providing for incentive payments to meaningful EHR users but not after 2016).

²¹² *Id.* § 1395w-4(o)(2)(A)(i), (iii). The Secretary must publish the measures in advance in the Federal Register. *Id.* § 1395w-4(o)(2)(B)(i)(II). The Secretary must also “provide [a] preference” to clinical quality measures endorsed by the National Quality Forum. *Id.* § 1395w-4(o)(2)(B)(i)(I). The Forum provided a set of preliminary measures in 2011. See generally NAT’L QUALITY FORUM, IDENTIFICATION OF POTENTIAL 2013 E-QUALITY MEASURES (2011), http://www.qualityforum.org/Publications/2011/02/Meaningful_Use_Final_Report.aspx [<https://perma.cc/SC3N-2Q3V>].

²¹³ 42 U.S.C. § 1395w-4(o)(2)(A)(iii).

to concern those services. In other words, the Secretary could impose payment penalties for not reporting non-CMS beneficiaries' information.²¹⁴

“[V]irtually all” hospitals²¹⁵ and over 90% of primary care providers accept Medicare or Medicaid and would be affected by such penalties.²¹⁶ As CMS has issued almost yearly updates on appropriate reporting measures and methodologies on meaningful use since the program's inception,²¹⁷ slowly incorporating the reporting of nonbeneficiary data is a viable approach.²¹⁸

All this data would be made available to the Patient Centered Outcomes Research Institute for secondary research under the ACA. Recall that CMS must “make available to the Institute such data collected . . . under” Medicare and Medicaid.²¹⁹ The data would be reported “under” an incentive program connected to Medicare and Medicaid.

As technology develops, we may consider alternative approaches to collecting data from all patients. One potential method is an interactive sampling approach in which the CMS data registry program looks at the patient's profile and only accepts the data of those who are underrepresented in the pool.²²⁰ Similarly, under a distributed architecture model, which I discuss below,²²¹ research would be done remotely, with no data collection.

²¹⁴ CMS may have already realized that it has this authority. Under the meaningful use program as it is currently constituted, providers must report certain clinical quality measures (CQM). *Id.* These are usually aggregate calculations based on *all* of the patients the doctors have seen. Thus, if the CQM is (number of minors taking certain medication) over (total number of minors), the provider must include *all* minors in her practice, not just those receiving Medicaid. As meaningful use transitions from seeking aggregate data reporting to transmission of granular EHRs, CMS should make data available, not just of the recipients of public benefit recipients, but of all individuals in the practice. As I explain below, this need not require CMS to collect the data itself. *See infra* note 286 and accompanying text.

²¹⁵ Senate Rules Comm., Bill Analysis: SB 233, Third Reading (Apr. 27, 2011), http://www.leginfo.ca.gov/pub/1112/bill/sen/sb_02010250/sb_233_cfa_20110518_160126_sen_floor.html [https://perma.cc/S2S7-X95B].

²¹⁶ Cristina Boccuti et al., *Primary Care Physicians Accepting Medicare: A Snapshot*, HENRY J. KAISER FAM. FOUND. (Oct. 30, 2015), <http://kff.org/medicare/issue-brief/primary-care-physicians-accepting-medicare-a-snapshot/> [https://perma.cc/J87Z-67BU].

²¹⁷ *See, e.g., 2016 Program Requirements*, CMS.GOV, <https://www.cms.gov/regulations-and-guidance/legislation/ehrincentiveprograms/2016programrequirements.html> [https://perma.cc/D6U4-K64L] (last updated Feb. 14, 2017).

²¹⁸ What is key, of course, is for CMS to develop “the capacity to accept the information electronically” as the statute requires. 42 U.S.C. § 1395w-4(o)(2)(B)(ii).

²¹⁹ *Id.* § 1320e(d)(3)(A).

²²⁰ *See generally* Mansurul Bhuiyan et al., *Interactive Pattern Mining on Hidden Data: A Sampling-Based Solution*, in CIKM '12, THE PROCEEDINGS OF THE 21ST ACM INTERNATIONAL CONFERENCE ON INFORMATION AND KNOWLEDGE MANAGEMENT 95 (2012), <http://dl.acm.org/citation.cfm?id=2396777> [https://perma.cc/VK8H-H9SK] (describing the process and benefits of interactive sampling).

²²¹ *See infra* notes 240, 286 and accompanying text.

2. Claims Data

Although EHR data collected through the CMS program will satisfy the needs of most secondary clinical research, claims data also has additional value for four reasons. First, and most importantly, under existing medical technology, the meaningful use program is probably a few years away from reasonably demanding that doctors report individual-level clinical measures. Claims data, however, has long been and will continue to be transmitted. Second, claims data contains information that EHR data lacks that is vital for health delivery planning, such as the price of services. Third, claims data goes back many years and therefore has “longitudinal depth.”²²² Fourth, claims data can be useful as a cross check of EHR data to minimize data error. Hence, some secondary researchers argue that “neither claims data nor [EHR] data alone can provide a complete, accurate, and timely view of a person’s health status.”²²³

Since the meaningful use program only gives CMS the authority to require meaningful use of EHR data, it lacks the authority to demand claims data. However, states have traditionally regulated insurance and created all-payer claims databases, as described above.²²⁴ The federal government has so far not been involved in this effort in any substantial way.²²⁵

After a recent decision by the Supreme Court, this must change. In *Gobeille*, the Court held that ERISA preempted states from requiring self-funded plans to report data.²²⁶ In explaining why, the Court relied on the broad authority of the Department of Labor to require the reporting of information from self-insured plans.²²⁷

²²² *Claims Data: The Good, the Bad and the Ugly*, FLOW HEALTH (May 16, 2014), <http://blog.flowhealth.com/claims-data-the-good-the-bad-and-the-ugly/> [<https://perma.cc/6HD8-HUAT>].

²²³ WILSON & BOCK, *supra* note 25, at 4.

²²⁴ See *supra* text accompanying notes 62–64.

²²⁵ But it has provided grants to build APCDs. See *The Center for Consumer Information & Insurance Oversight: Rate Review Grants*, CMS.GOV, <https://www.cms.gov/CCIIO/Resources/Rate-Review-Grants/index.html> [<https://perma.cc/T55F-GU7V>] (describing the \$250 million available to states for reviewing health care costs, some of which goes to fund data centers).

²²⁶ *Gobeille v. Liberty Mut. Ins. Co.*, 136 S. Ct. 936, 947 (2016); see also *supra* text accompanying notes 65–67.

²²⁷ The Court details this authority in the following passage:

The Secretary also may, “in connection” with any research, “collect, compile, analyze, and publish data, information, and statistics relating to” plans. [42 U.S.C.] § 1143(a)(1); see also § 1143(a)(3) (approving “other studies relating to employee benefit plans, the matters regulated by this subchapter, and the enforcement procedures provided for under this subchapter”).

ERISA further permits the Secretary of Labor to “requir[e] any information or data from any [plan] where he finds such data or information is necessary to carry out the purposes of” the statute, § 1024(a)(2)(B) The Secretary has the general power to promulgate regulations “necessary or appropriate” to administer the statute, § 1135

Gobeille, 136 S. Ct. at 944 (alteration in original) (italics added).

So far, the Department of Labor has primarily sought information about plan financing.²²⁸ However, as the Court discussed,²²⁹ the Department of Labor can now work with the states (and HHS) to create a uniform system of insurance information reporting.²³⁰ And with the Court's implicit blessing, such a plan is unlikely to face serious legal challenges, although other problems would remain.²³¹

Nonetheless, coordination will be required. A robust database requires claims data from both self-insured and other plans, which is impossible for one government level to achieve. ERISA prevents the Department of Labor from devolving all authority to collect self-insured data to states, as the Court was adamant that Congress did not intend for states to collect such data.²³² On the other hand, federalism and administrative authority concerns will prevent federal agencies from collecting data from all plans: ERISA gives the Department of Labor authority over only self-insured plans and the ACA gives HHS only limited data collection authority.²³³ A cooperative, voluntary federal–state venture would bring together the data that falls within the respective competencies of federal and state governments in a single data pool.²³⁴

²²⁸ See *id.* at 954 (Ginsburg, J., dissenting) (“ERISA-covered benefit plans must . . . file annual reports containing financial and actuarial data to . . . the Secretary of Labor . . .”).

²²⁹ See *id.* at 944 (majority opinion) (explaining that ERISA oversight systems “are intended to be uniform” across all fifty states); see also *id.* at 949 (Breyer, J., concurring) (noting that the grant of authority allows state–federal cooperation).

²³⁰ Bill Sage also appears to suggest that *Gobeille* is a net positive for creating collaboration between states and the federal government. William Sage, *Out of Many, One: ERISA Preemption, State All-Payer Claims Database Laws, and the Goals of Transparency*, HEALTH AFF. BLOG (Mar. 10, 2016), <http://healthaffairs.org/blog/2016/03/10/out-of-many-one-erisa-preemption-state-all-payer-claims-database-laws-and-the-goals-of-transparency/> [<https://perma.cc/VWZ2-CERR>].

²³¹ For example, information collected by the Secretary is “public information,” except certain kinds of beneficiary information. 29 U.S.C. § 1026(a) (2012).

The Secretary “shall make” such information available in various contexts. *Id.* However, to avoid constitutional problems, this requirement should, like CMS data, be read as subject to certain privacy protections and subject to at least the same procedural collection requirements as CMS data. For additional suggestions, see generally Bland et al., *supra* note 63, which discusses nonstate APCD options for further data collection.

²³² See *Gobeille*, 136 S. Ct. at 945 (majority opinion) (“[T]he uniform rule design of ERISA makes it clear that [decisions concerning the exemption self-insured plans] are for federal authorities, not for the separate States.”).

²³³ See 42 U.S.C. § 18031(e)(3)(A) (2012) (listing the type of data that health plans must submit to the Secretary). Notably, the Department of Labor has relied on this specific authority in its rulemaking, rather than the more general authority cited in *Gobeille*. See 136 S. Ct. at 944 (describing the Secretary's broad grant of authority to establish additional reporting and disclosure requirements for ERISA plans).

²³⁴ Such a scheme would not be unprecedented. For example, “[t]he National (Nationwide) Inpatient Sample . . . is the largest publicly available all-payer inpatient health care database in the United States,” and was “[d]eveloped through a Federal-State-Industry partnership sponsored by the Agency for Healthcare Research and Quality.” *Overview of the National (Nationwide) Inpatient Sample*, H-CUP, <https://www.hcup-us.ahrq.gov/nisoverview.jsp> [<https://perma.cc/YVU6-4G7U>].

3. The HIPAA Exclusion

Some individuals will pay out of pocket, and their EHR and claims data may never be included in the medical record as I explain above.²³⁵ Changing this requirement would require amending the 2009 statute that put it into place.²³⁶ Nonetheless, the ability to exclude information from health records does serve some important purposes, and changes should be implemented carefully. Preventing providers from transmitting certain information to payers (and effectively, from including the information in the health record) when the patient pays out of pocket can assist vulnerable patients. For example, dependent children or spouses may want to conceal from the primary policyholder (or anyone else with access) that they had sought birth control or STD tests and may therefore pay out of pocket for the service.²³⁷ Thus, while the out-of-pocket data exemption may allow the rich to exempt themselves from the data pool, it also allows certain individuals to take care of their health, avoid domestic violence, and exercise their autonomy. A nuanced statutory response is therefore necessary.

One possible solution is to require a provider to comply with the data exclusion request only if the plan includes more than one member. In other cases, providers could be required to collect and enter health information. This allows vulnerable individuals to withhold data from the policyholder but would include the remaining data within the secondary research data pool.

4. Data Collection from the Healthy

As I explain above,²³⁸ healthier individuals escape information burdens even though their information can be extremely useful. Incentives should be introduced to encourage checkups that would both improve preventative medicine and allow for systematic health information collection from the healthy. In the meantime, data from wellness programs can help close the gap.

First, we could institute tax incentives for preventative care and checkups for healthy individuals. Just as individuals must now generally carry health insurance, individuals would be required to go to the doctor to ensure that any diseases can be detected early and addressed more effectively at the least cost to society, and

²³⁵ 45 C.F.R. § 164.522(a)(1)(vi) (2013); *see also supra* text accompanying note 69.

²³⁶ *See* 42 U.S.C. § 17935(a) (outlining when entities must comply with an individual's request to restrict the disclosure of protected health information under 45 C.F.R. § 164.522).

²³⁷ *See* Laura Covarrubias, *The Girls' Guide to Getting Some Privacy on Your Parents' Health Insurance*, BEDSIDER (May 16, 2013), <https://bedsider.org/features/275-the-girls-guide-to-getting-some-privacy-on-your-parents-health-insurance> [<https://perma.cc/2X52-JLTK>] (providing tips for how to conceal treatments from parents while remaining on their health insurance).

²³⁸ *See supra* text accompanying notes 75–78.

that their information can be used to benefit society as a whole.²³⁹ States might also mandate such checkups within their police power as they do vaccines. Other countries, such as Japan, are close to having similar requirements.²⁴⁰

Until the point where such a program can be enacted, we might rely on wellness programs, where *employers*, rather than states, incentivize individuals to get checkups. These programs, which the ACA encourages, often offer incentives in the form of premium discounts or gift cards for participating employees of up to 30% of the premium amount.²⁴¹ Incentives may also be tied to certain biometric outcomes that require employees to be tested for cholesterol, blood pressure, or body mass index. Some programs also allow testing for nicotine exposure.²⁴²

There are many problems with wellness programs beyond the scope of this Article.²⁴³ However, for our purposes, what is key is that data from wellness programs will never become part of either EHR or claims reporting.

²³⁹ At least one major presidential primary candidate has supported this approach in the past. See Amy Lorentzen, *Edwards Backs Mandatory Preventive Care*, WASH. POST (Sept. 2, 2007, 8:06 PM), <http://www.washingtonpost.com/wp-dyn/content/article/2007/09/02/AR2007090200743.html> [<https://perma.cc/4S6N-9BW6>] (reporting that in 2007, John Edwards proposed a mandatory preventative care measure as part of his universal health care plan). Others have promoted similar ideas more recently. See, e.g., Dan Mangan, *Take an 'Un Sick Day': Paid Time Off for Your Health-Care Checkups*, CNBC (Oct. 11, 2016, 12:01 AM), <https://www.cnbc.com/2016/10/10/take-an-unsick-day-paid-time-off-for-your-health-care-checkups.html?source=twitter> [<https://perma.cc/F6DP-DZPF>] (proposing a mandatory paid day off solely for the use of annual checkups and other preventive care). This approach, they argue, will reduce costs on the medical system and save lives.

²⁴⁰ See Mami Maruko, *Simple Tests Fill Health-Check Gaps*, JAPAN TIMES (Dec. 30, 2013), <http://www.japantimes.co.jp/news/2013/12/30/reference/simple-tests-fill-health-check-gaps/#.WSO hFJOzJ3I> [<https://perma.cc/AZ56-5TLX>] (“[H]ealth checkups are often mandatory for corporate or institutional employees . . .”); see also Tom Lomax, *What to Expect when Undergoing a Medical Check-Up in Japan*, JAPAN TODAY (May 27, 2013, 6:29 AM), <https://japantoday.com/category/features/health/what-to-expect-when-undergoing-a-medical-check-up-in-japan> [<https://perma.cc/F9V H-YTQK>] (“[R]egulations fall short of making it compulsory for employees to have an annual medical. However, government incentives are in place in larger firms to ensure that a high percentage of staff complete their check-up and file the report with their firm. This can lead to staff being pursued to have a medical even if they don’t want one.”).

²⁴¹ See Incentives for Nondiscriminatory Wellness Programs in Group Health Plans, 78 Fed. Reg. 33,158, 33,158 (June 3, 2013) (allowing premium discounts “in return for adherence to certain programs of health promotion and disease prevention”). The Affordable Care Act amended the HIPAA nondiscrimination and wellness provisions. Patient Protection and Affordable Care Act, Pub. L. No. 111-148, § 1201, 124 Stat. 119, 154 (2010) (codified as amended in scattered sections of the U.S. Code).

²⁴² See Kristin Madison, *The ACA, the ADA, and Wellness Program Incentives*, HEALTH AFF. BLOG (May 13, 2015), <https://healthaffairs.org/blog/2015/05/13/the-aca-the-ada-and-wellness-program-incentives/> [<https://perma.cc/Y9T5-Y9LU>] (providing an overview of wellness programs and the respective federal regulation).

²⁴³ For a good overview of the problems, see generally *Special Issue: The Law and Politics of Workplace Wellness*, 39 J. HEALTH POL. POL’Y & L. 955 (2014).

Although there is scant information about these programs,²⁴⁴ they can easily serve as data siphons for healthy individuals. Healthy individuals with jobs, who would otherwise have gone to doctors for annual checkups, can now confirm their own good health without ever contributing information to the data pool. Combining data from these programs into EHRs will create efficiencies, improve data protection, and address inequities in data collection and use.

For wellness data to be collected, federal entities would likely have to intervene once more. After *Gobeille*, states are likely preempted from collecting such data. However, the ACA can be read to provide such authority to HHS and Labor. For example, it allows the departments to collect data for reports, and it gives them authority to promulgate additional regulations “in connection” with wellness programs.²⁴⁵

5. Other Data Collection Policies

Apart from slight alterations to existing regulatory and legal regimes, it may be possible to put into place stronger collection measures to ensure data collection from *all* providers, not just the vast majority that accept Medicare and Medicaid. This more intrusive approach would extend reporting requirements even further.

States are best poised to impose this requirement. As the respondent observed during *Gobeille* oral arguments, instead of having self-insured plans report to all-payer claims databases, states could require providers to report claims data to the state.²⁴⁶ This could extend to the reporting of EHR information as well. The federal government could then coordinate with the states to ensure uniform reporting.

Most states and the federal government currently lack the infrastructure and funding to establish such a program, and the incremental CMS-led program that I have proposed is more practical for now. However, several states are developing “health information exchanges.”²⁴⁷ Most of these entities focus only on data exchange between providers for treatment purposes. As the schemes progress, they should permit remote queries for clinical research purposes

²⁴⁴ See Madison, *supra* note 242 (“[H]igh-quality evidence on the impact of employer wellness incentives is in short supply.”).

²⁴⁵ See 42 U.S.C. § 300gg-4(n) (2012) (“Nothing in this section shall be construed as prohibiting the Secretaries of Labor, Health and Human Services, or the Treasury from promulgating regulations in connection with this section.”); see also *id.* § 300gg-4(m)(2) (permitting collection of additional data from employers who provide employees with access to wellness programs).

²⁴⁶ Transcript of Oral Argument at 44, *Gobeille v. Liberty Mut. Ins. Co.*, 136 S. Ct. 936 (2016) (No. 14-181).

²⁴⁷ See generally PRASHILA DULLABH ET AL., NORC AT THE UNIV. OF CHI., EVALUATION OF THE STATE HIE COOPERATIVE AGREEMENT PROGRAM (2016).

as well.²⁴⁸ And the federal government should continue recently introduced efforts to assist such exchanges.²⁴⁹

Congress could also probably impose collection requirements on large insurers and providers under its commerce power. It could also nudge states by providing limited incentives to impose such requirements over time.²⁵⁰ Finally, it has also worked directly with providers to encourage data sharing. The Institute, for example, has set up links with numerous private networks to collect data,²⁵¹ and President Obama's Precision Medicine Initiative seeks to do the same.²⁵² These programs' guiding bodies must carefully assess which entities to include within these networks as part of a strategic plan to ensure that information collection burdens are widely distributed population-wide.

B. Research

This Article is focused primarily on the burdens of information collection, and changes in that process should be the first step toward equity. However, once collected, the data is then passed on to other entities for analysis. Each transfer or query of health information increases the risk of breach and the harms of commodification. There may be data breaches at the time of transmission, but also within research facilities.

We must, at the very least, ensure that risk is equally distributed across groups. We can do so in various ways. For example, individuals whose data has been involved in multiple studies may suffer a higher risk of reidentification or breach than others, so their data should be exempted at some point if possible.²⁵³ We must also introduce steps to assess the background of the patient

²⁴⁸ I should note that the new version of the Common Rule will only require institutional review board (IRB) approval from one IRB per research project in certain cases. Federal Policy for the Protection of Human Subjects, 82 Fed. Reg. 7149, 7208 (Jan. 19, 2017). Thus, a single approval would be sufficient for a researcher to get access to data statewide for a specific project.

²⁴⁹ The federal government has provided assistance to these exchanges consistently. Most recently, however, it has declined to mandate such exchanges: "[W]e do not believe it is appropriate for us to require or mandate this option, as states may have various options or paths to increase EHR and HIE adoption outside of their managed care contracts." Medicaid and Children's Health Insurance Program (CHIP) Programs; Medicaid Managed Care, CHIP Delivered in Managed Care, and Revisions Related to Third Party Liability, 81 Fed. Reg. 27,497, 27,649 (May 6, 2016).

²⁵⁰ *Cf.* Nat'l Fed'n of Indep. Bus. v. Sebelius, 132 S. Ct. 2566, 2601 (2012) (delineating the limits on "Congress's power under the Spending Clause to secure state compliance with federal objectives").

²⁵¹ *See supra* note 44.

²⁵² THE PRECISION MEDICINE INITIATIVE, *supra* note 6, at 40.

²⁵³ *See* Comment Letter from Lee Tien, Senior Staff Attorney, Elec. Frontier Found., to Dep't of Health & Human Servs. Regarding the Notice of Proposed Rulemaking: Human Subjects Research Protections: Enhancing Protections for Research Subjects and Reducing Burden, Delay, and Ambiguity for Investigators, Docket ID number HHS-OPHS-2015-0008, at 8 (Jan. 6, 2015), https://www.eff.org/files/2016/01/07/eff_common_rule_nprm_comment.pdf [<https://perma.cc/GB7T-WFPA>] (proposing such an exemption).

whose data is used to make sure that the burdens are spread across various social groups. At first, such an assessment would be based on a rough calculus that looks at how the patient paid for care and other demographic information. As time goes on and more data is collected in medical records, including social determinants of health, other, more exact parameters will include income information and other measures of welfare, ranging from age to family size or disability, as in the taxation context.²⁵⁴

As I explain in Section III.B, however, some may argue that we should impose higher burdens on the better off.²⁵⁵ One might argue that all else being equal, we should use the data of individuals with less hardship for research.²⁵⁶

For example, research seeking to identify the genetic determinants of a specific disease may not require representation from all socioeconomic groups. And as a general matter, depending on the study and the relevant statistical method, only a certain number of records need to be included to yield the appropriate statistical confidence.²⁵⁷ In cases where selection bias is minimal, we might focus information burdens on the richer and younger by prioritizing their information for transmission to researchers. Information coming from Medicare or Medicaid beneficiaries should be exempt to a greater degree.

²⁵⁴ CMS has the ability to cross-reference data with beneficiary financial information. Medicare Parts B and D have income sensitive premiums. Juliette Cubanski & Tricia Neuman, *Medicare's Income-Related Premiums: A Data Note*, HENRY J. KAISER FAM. FOUND. (July 3, 2015), <http://kff.org/medicare/issue-brief/medicares-income-related-premiums-a-data-note/> [<https://perma.cc/A4K2-MK94>]; see also *Means-Testing of the Medicare Part B Premium*, NAT'L COMMITTEE TO PRESERVE SOC. SECURITY & MEDICARE (July 2009), <http://www.ncpssm.org/Document/ArticleID/777> [<https://perma.cc/EG9F-GLM7>] (“[T]he Social Security Administration (SSA) automatically determines the amount of the Medicare Part B premium using tax records filed with the Internal Revenue Service two or three years previously.”).

²⁵⁵ See Belmont Report, *supra* note 12, at 23, 196-97 (“Some populations, especially institutionalized ones, are already burdened in many ways by their infirmities and environments. When research is proposed that involves risks and does not include a therapeutic component, other less burdened classes of persons should be called upon first to accept these risks of research.”).

²⁵⁶ Those who reject such graduated risk distribution and argue that risk should be equally distributed across all individuals would reject this scheme.

²⁵⁷ See OFFICE OF THE NAT'L COORDINATOR FOR HEALTH INFO. TECH., DEP'T OF HEALTH & HUMAN SERVS., REPORT TO CONGRESS: UPDATE ON THE ADOPTION OF HEALTH INFORMATION TECHNOLOGY AND RELATED EFFORTS TO FACILITATE THE ELECTRONIC USE AND EXCHANGE OF HEALTH INFORMATION 21-22 (2014), http://www.healthit.gov/sites/default/files/rhc_adoption_and_exchange9302014.pdf [<https://perma.cc/7P5T-S2HA>] (describing the Comprehensive Primary Care initiative, which “[i]n an effort to reduce providers' quality reporting burden, . . . leverages the EHR Incentive Programs to measure quality by using a subset of the clinical quality measures specified for reporting”); see also THE PRECISION MEDICINE INITIATIVE, *supra* note 6, at 22 (explaining statistical power requirements in nested unmatched case-control design, which limit the number of people required to achieve desired levels of confidence in case control studies).

Information coming from self-paid patients, or those in higher priced insurance plans, may be subject to greater use.²⁵⁸

There are of course limits to this goal of equitable adjustment. Analyses involving, for example, social determinants of health, are often biased based on population. In those cases, we must seek and use the information of those who are worse off to make sure that they remain represented in the data.

When researchers seek data, IRB protocols should take into account equity principles and consider how the data request can be structured to distribute burdens equitably. CMS and its contractors can continue to review research requests, but in doing so, should also consider and discuss equity-based considerations with researchers.

C. Coordination

Perhaps the biggest problem that faces the modern health information effort is that policymakers and scholars understand it through the traditional lens of health information collection during public health emergencies. Traditional collection does not comprise a single system. Each act of collection is justified by an immediate purpose—stopping a particular outbreak or running a specific trial. Each program is therefore “ad hoc . . . [:] the laws have developed by putting out fires, without comprehensive planning for modern public health problems.”²⁵⁹

Because of this, policy and academic approaches also treat each act of collection as a standalone enterprise.²⁶⁰ Considerations advanced with respect to one health collection project, such as increasing the number of providers that are part of CMS’s meaningful use program, may prove completely inapplicable with respect to other health information projects such as the FDA’s post-market drug surveillance. Each subagency has collected information for agency specific projects: the FDA, for example, collects prescription data,

²⁵⁸ See *Health Plan Categories*, HEALTHCARE.GOV, <https://www.healthcare.gov/glossary/health-plan-categories/> [<https://perma.cc/Y6US-5SQ6>].

²⁵⁹ Lawrence O. Gostin, *The Future of Public Health Law*, 12 AM. J.L. & MED. 461, 476 (1986).

²⁶⁰ A small group of scholars arguably provide a systematic approach by suggesting the creation of a national information market where trusted intermediaries can hold and sell the information—recouping gains for the benefits of the patients and perhaps their providers. See Mark A. Hall, *Property, Privacy, and the Pursuit of Interconnected Electronic Medical Records*, 95 IOWA L. REV. 631, 638 (2010) (arguing that “patients [should be allowed] to license limited rights to their medical information with trusted and regulated intermediaries”); see also ALAN F. WESTIN, *PRIVACY AND FREEDOM* 324-25 (1967) (“[Personal information . . . should be defined as a property right, with all the restraints on interference by public or private authorities and due-process guarantees [of] our law of property . . .”). But there is no indication that policy is moving in that direction, nor any guarantee that this will increase information access. Further, it would be nearly impossible to create a reliable and unbiased health information source with this approach.

CMS collects Medicare data, and so on.²⁶¹ This may, in some cases, lead to waste and duplication—one can imagine multiple agencies accessing data from the same doctors regarding the same patients. This has prevented us from developing any systemwide principles or goals with which to evaluate health collection efforts at a more general level.

The changes I recommend will require some form of coordination among government entities, resulting in the formation of a single entity or task force. This entity must oversee data coordination efforts through CMS, FDA, the Department of Labor, state claims databases, wearable devices, wellness programs, and consumer health initiatives. It must then ensure that data is accessed for research in ways that are fair by keeping track of all research projects. Where possible, the data of the same individuals or groups should not be used in a multitude of projects: risks should be evenly distributed. In addition to ensuring fairness, the entity could also play a role in curating data quality by coordinating data collection formats and avoiding data redundancy.

This entity could also develop an optimum multi-database computational approach, though designers may also need to consider latency costs and duplication.²⁶² As it currently stands, researchers may have to collect data from multiple sources and manipulate it in various ways to ensure that the metrics employed are congruent. With appropriate technology and standards, data need never be collected: all research queries could be done remotely.

²⁶¹ Medicare Part D data, that is, prescription data, might conceivably overlap with the data that the FDA collects through its Sentinel program which carries out post market surveillance of drug incidents. *See infra* note 262.

²⁶² Carol Diamond explains that in such a model, “personally identified information is held only at the source, and data are cleaned and analyzed in a common way . . . before being sent in a standardized format.” Carol C. Diamond et al., *Collecting and Sharing Data for Population Health: A New Paradigm*, 28 HEALTH AFF. 454, 459 (2009). Diamond, however, warns of a potential latency effect. *See id.* at 458 (“[M]any forms of analysis are significantly delayed, including those that should ideally be close to real time . . .”). Distributed approaches may be fine for long-term tracking systems like the FDA’s mini-Sentinel system, *see generally* FDA, MINI-SENTINEL: PRINCIPLES AND POLICIES (2013) (describing the Sentinel Initiative, a “long-term, multifaceted effort to create a national electronic system for monitoring the safety of FDA-regulated medical products”), but cannot work for real-time public health interventions. There may also be duplication issues without centralized storage. *See* CLAUDIA GROSSMANN ET AL., INST. OF MED. OF THE NAT’L ACADS., CLINICAL DATA AS THE BASIC STAPLE FOR A LEARNING HEALTH SYSTEM 114-15, 150-51 (2010) (noting that the “chief disadvantage of [a distributive research network] is the patient deduplication issue”). Data lake approaches may offer a new alternative. *See* Kerry Raminiak, *Data Lakes and the Promise of Unsiloes Data*, PWC (May 6, 2015), <http://www.pwc.com/us/en/technology-forecast/2014/cloud-computing/features/data-lakes.html> [<https://perma.cc/M24A-DQYG>] (describing how Data lakes allow “disparate records to be stored in their native formats for later parsing, rather than forcing all-or-nothing integration up front as in a data warehousing scenario”). Another alternative is a system that collects core data elements in a central location with the remaining elements distributed, as envisaged in the Precision Medicine Initiative. THE PRECISION MEDICINE INITIATIVE, *supra* note 6, at 37.

There are three potential candidates for the coordinating role. The Patient Centered Outcomes Research Institute has been given the responsibility of developing, disseminating, prioritizing, and contracting for research to “assist patients, clinicians, purchasers, and policy-makers in making informed health decisions.”²⁶³ It has the authority to “identify national priorities for research” through cost–benefit analysis.²⁶⁴ It may “also request and obtain data from Federal, State, or private entities, including data from clinical databases and registries.”²⁶⁵ The Institute itself is not a government entity,²⁶⁶ but is instead administered by a board of representatives from numerous private and public organizations.²⁶⁷ It is therefore well-positioned to work with all stakeholders in collecting data and disseminating research.

Nonetheless, this strength might also be a weakness. For the Institute to create data collection initiatives that carry the force of law, it would need to become a government entity. Cooperation with other agencies might be undermined if private citizens sit on its board. Furthermore, the Institute is supported by a trust fund without additional appropriation.²⁶⁸ The amount in the trust fund is insufficient to support incentive programs for data collection such as meaningful use.

Next, the Agency for Healthcare Research and Quality is a subagency within HHS. One of the Agency’s purposes is to “enhance . . . health services” through a range of health-delivery and clinical research,²⁶⁹ provide research grants,²⁷⁰ and issue clinical guidelines.²⁷¹ In particular, it is tasked with several data-focused objectives. It must assess “the utility and comparability of health information data and medical vocabularies by addressing issues related to the content, structure, definitions and coding of such information,” and must promote “the use of computer-based health records in all settings.”²⁷² Further, the Secretary, acting through the Director, “shall provide for the coordination of relevant Federal health programs . . . including the development . . . of . . . health outcomes research data networks, in order to

²⁶³ 42 U.S.C. § 1320e(c) (2012).

²⁶⁴ *Id.* § 1320e(d)(1)(A).

²⁶⁵ *Id.* § 1320e(d)(3)(A).

²⁶⁶ *See id.* § 1320e(b)(1) (noting that the Institute “is neither an agency nor establishment of the United States Government”).

²⁶⁷ *Id.* § 1320e(f) (outlining the representatives from each industry that must comprise the Board of Governors).

²⁶⁸ 42 U.S.C. § 1320e(b)(3).

²⁶⁹ *Id.* § 299(b).

²⁷⁰ *Id.* § 299(a).

²⁷¹ *See* FRIENDS OF AHRQ, AHRQ: 15 YEARS OF TRANSFORMING CARE AND IMPROVING HEALTH 8 (2014) (“[AHRQ] provides physicians . . . detailed information on evidence-based clinical practice guidelines.”).

²⁷² 42 U.S.C. § 299b-3(a)(5)–(6).

develop and maintain a comprehensive, interoperable data network to collect, link, and analyze data on outcomes and effectiveness from multiple sources, including electronic health records.”²⁷³

However, the Agency’s mandate is too broad. Other priorities include developing decision aids to assist patients and doctors with shared decisionmaking.²⁷⁴ Second, it lacks authority with respect to data collection. At most, it can expect “voluntary collaboration” from private entities.²⁷⁵ Third, it lacks authority even with respect to federal data networks: although it has the authority to issue data collection standards, “where [these] standards . . . may affect the administration of other [HHS] programs . . . , they shall be in the form of recommendations to the Secretary for such program.”²⁷⁶ Indeed, the ACA deliberately put into place the Institute,²⁷⁷ which replicates many of the Agency’s functions. The Agency is tasked with assisting the Institute in “dissemination of the Institute’s research findings.”²⁷⁸

The final candidate for coordinating the secondary research network is the Office of the National Coordinator for Health Information Technology. The Office was established in 2004, by executive order.²⁷⁹ The HITECH Act established this subagency by statute five years later to further the “development of a nationwide health information technology infrastructure that allows for the electronic use and exchange of information” in order to “improve[] health care quality, reduce[] medical errors, reduce[] health disparities, . . . advance[] the delivery of patient-centered medical care,” and “facilitate[] health and clinical research and health care quality.”²⁸⁰ The Coordinator is given the task of “updat[ing] the Federal Health IT Strategic Plan . . . to include specific objectives, milestones, and metrics.”²⁸¹

Yet, when it comes down to brass tacks, the Coordinator lacks the authority to do much beyond issue technical specifications. His primary tasks are to “review and determine whether to endorse [health information technology] standard[s], implementation specification[s], and certification criteri[a],” and to “review Federal health information technology investments.”²⁸² Incentive program money is not siphoned through the Office. Rather, the Secretary

²⁷³ *Id.* § 299b-37(f).

²⁷⁴ *Id.* § 299b-36.

²⁷⁵ *Id.* 299b-7(a)(3)(A)(iv).

²⁷⁶ *Id.* § 299c-2(a)(2).

²⁷⁷ Patient Protection and Affordable Care Act, Pub. L. No. 111-148, § 1181(b)(1), 124 Stat. 119, 728 (2010) (codified as amended in scattered sections of the U.S. Code).

²⁷⁸ 42 U.S.C. § 299b-37(a)(2).

²⁷⁹ David J. Brailer, *Guiding the Health Information Technology Agenda*, 29 HEALTH AFF. 586, 586 (2010).

²⁸⁰ 42 U.S.C. § 300jj-11(b)(2),(8).

²⁸¹ *Id.* § 300jj-11(c)(3).

²⁸² *Id.* § 300jj-11(c)(1)(A),(C).

“invest[s] funds through the different agencies with expertise,”²⁸³ thus leaving authority with agencies outside the Office. Thus, real power remains with larger HHS agencies such as CMS and the FDA.

There are two possible solutions to achieving coordination. First, the Secretaries of HHS and Labor could create a task force with representatives from Labor, the Agency, and the Office of the National Coordinator, with the Institute playing an advisory role. The combined statutory authority I recite above would provide this task force with the authority to direct information collection prerogatives. Even if CMS distributes the incentive money, its role would be mainly ministerial—the task force, with the authority of the Secretary behind it, would determine how the money should be distributed.

Finally, Congress could combine the sometimes redundant functions of the various HHS agencies with respect to data collection and research into a single new agency. Only with this combined power could a coordinated research initiative be created that respects equity across data collection programs.

CONCLUSION

Health information collection carries great promise for society, but collection policies facilitate inequitable distribution of information burdens and risks. Although other forces may head off this inequity, ethical regulation would serve to distribute the burdens equitably across society to minimize the encumbrances on the poor and elderly.

There are, of course, limits to the goal of equity. Sometimes equity will not be technically *feasible* if we are to help certain underserved populations. We must seek and use the information of those who are poor and elderly in order to assist them. Indeed, while this Article focuses on the *burdens* of health information collection and rendering them equitable, fairness also would suggest that the *benefits* of collection and research should go to those who are worse off, which is not always the case.²⁸⁴

Some may also object to the organizational approach I offer, which appears to focus on government coordination, because of concerns arising from security and political pragmatism. Yet, the government may be as well or better placed to address security risk. As I note above, data can be stored in multiple locations and can be linked for specific queries.²⁸⁵ The government never holds on to the data itself.²⁸⁶ At the same time, a centralized government

²⁸³ *Id.* § 300jj-31(a).

²⁸⁴ *See supra* text accompanying notes 159–61.

²⁸⁵ *See supra* note 262 and accompanying text.

²⁸⁶ *See, e.g.,* FDA, *supra* note 262, at 2 (“Mini-Sentinel uses a distributed data approach in which Data Partners retain control over data in their possession as a result of normal activities.”). Most breaches occur due to human error. Private health care providers “experience frequent staff

point of access can help us ensure that “only bona fide researchers can obtain access . . . to preserve privacy and confidentiality.”²⁸⁷

These changes will not come about in a straightforward way. Equity intersects with other values—autonomy in individual encounters, consent requirements, and political and technological feasibility, to name a few. As conversations progress, different visions of autonomy and solidarity, and beneficence and benefit, will be tested. As these visions are deployed in administrative and legislative crucibles, compromises, piecemeal, and pluralistic policies will take shape. Data collection efforts will be stalled and delayed. But in general, the result will be a system that best conforms to our deepest ideals of justice in the distribution of obligation.

Stepping back to consider the ethics of health information collection at the systemic level—rather than at that of individual encounters—reveals it to be a site where visions of citizenship play out. Individuals and the community mutually constitute and assist each other through this information. Bioethics demands, and law must create, a system of just obligations that bind both individuals and society to each other at the site of health information collection.

turnover, which results in a continual challenge of adequately educating” the staff. GROSSMAN ET AL., *supra* note 262, at 195. Training workers in a single organization may be far easier than ensuring that multiple entities meet training requirements. Further, providers have experienced breaches that have been comparable to or worse than those of the government. Compare Sisi Wei & Charles Ornstein, *Over 1,400 Health Data Breaches, but Few Fines*, PROPUBLICA (Feb. 27, 2015), <https://projects.propublica.org/graphics/healthcare-data-breaches> [<https://perma.cc/9WYM-TMLH>] (documenting 1419 large-scale data breaches within health care organizations and their business partners from 2009 through 2015), with A. Michael Froomkin, *Government Data Breaches*, 24 BERKELEY TECH. L.J. 1019, 1027 (2009) (citing one study which documented “110 breaches of state . . . federal, and military databases in 2008”). Indeed, the rate at which government entities suffer breaches has declined. *See id.* at 1027 n.35 (noting that there has been a fifty percent drop in reported breaches of U.S. government data from 2006 through 2009). Further, “those who aggregate and mine [big] data neither view their informational assets as public goods held on trust nor seem particularly interested in protecting the privacy of their data subjects.” Terry, *Protecting Patient Privacy*, *supra* note 11, at 389. Indeed, “[t]he truth lies in the opposite because the big data business model is selling information about their data subjects.” *Id.* Some others contend that the big data business model is even using data to discriminate. *See, e.g.*, Pasquale, *supra* note 163, at 108 n.43 (“There are also legitimate worries about discriminatory uses of information either not covered by existent privacy or anti-discrimination laws, or undetectable by workers.”).

²⁸⁷ Donna M. Gitter, *The Challenge of Achieving Open Source Sharing of Biobank Data*, in COMPARATIVE ISSUES IN THE GOVERNANCE OF RESEARCH BIOBANKS 165, 185 (Giovanni Pascuzzi et al. eds., 2013). Many suggest that smaller, decentralized data holding might be less secure: one argument is that “no individual company will ever have the level of security and keep up with the arms race” like bigger, more centralized repositories. Lucas Mearian, *Hackers Are Coming for Your Healthcare Records—Here’s Why*, CSO ONLINE (July 1, 2016, 4:24 AM), <http://www.csoonline.com/article/3090553/security/hackers-are-coming-for-your-healthcare-records-heres-why.html> [<https://perma.cc/M73A-N5N7>].