Draft copy, forthcoming 164 U. Pa. L. Rev. __ (2017).

Governing Health Information

Craig Konnoth*

In the not too distant future, the health information of all Americans will be collected as part of routine medical practice. The vast quantities of data thus collected will be combined to study correlations between medical conditions or treatments and genetic or behavioral profiles. Recent federal legislation and regulations have sought to make such collection universal and routine to achieve a broad set of public goals. The question we face today, then, is not whether we should incur privacy risk, but how we should distribute privacy risks in an equitable and legitimate manner. But so far, decisions regarding data collection focus only on the desirability of particular research questions, technological efficacy, and value maximization. This has resulted in an inequitable and illegitimate distribution of privacy risks, borne most by the most vulnerable in society.

To conceptualize the new health information collection system, and work out the principles that should guide it, this Article looks to the income taxation system. As a descriptive matter, taxation also involves routine and universal pooling of resources from all citizens for public goals. It is supported by similar normative accounts of how the individual owes duties to society and society to the individual, and indeed, how individual and society are mutually constituted. The analogy helps conceptualize the health information collection apparatus as a single system. Like the tax collection system, health information collection must engage with three key values: equitable distribution of burdens and benefits, legitimate and transparent processes, and efficient collection. The tax analogy helps guide health information institutions toward short- and long-term goals. By offering a well and widely understood, but at the same time, provocative, model, the analogy will promote discussion, and engagement in a system that has so far remained invisible to the broader public.

---

* Sharswood Fellow and Lecturer in Law, University of Pennsylvania Law School. J.D., Yale Law School; M.Phil., University of Cambridge. My thanks to Anita Allen, Regina Austin, Anne Barnhill, Stephanos Bibas, Glenn Cohen, Jonah Gelbach, Daniel Hemel, Allison Hoffman, Elizabeth Pendo, Bill Sage, Lindsay Wiley, Tess Wilkinson-Ryan, and participants of the William & Mary School of Law Workshop and the St. Louis University School of Law Health Scholars Workshop. Thanks are owed to Andrew Dean, Taylor Goodspeed, Rick Mula, and Emily Turner for research assistance, Alvin Dong for library assistance, and Benjamin Meltzer for formatting assistance.

Introduction 3
I. The New Model of Health Collection 12
A. Descriptive Characteristics 13
B. Normative Characteristics 14
II. The Case for a Taxation Analogy 16
A. The Descriptive Case 18
1. Burdens 18
2. Benefits 21
3. Government Intervention 22
B. The Normative Case 23
1. The Benefit Principle 24
2. The Redistribution Principle 27
3. The Solidarity Principle 28
4. The Regulatory Principle 31
III. Conceptualizing and Ordering the System 32
A. Systematic Perspective 34
B. Normative Principles 36
1. Substantive Justice: Fairness 37
2. Procedural Justice: Democratic Decisionmaking 44
a) Contrasting models 44
b) Analogous Ethical Grounding 46
c) Solving problems 47
3. Efficiency: Ease and Minimizing Waste 53
a) Correcting Inefficiencies 54
b) Considerations Against Government Leadership 58
Conclusion 60

INTRODUCTION

Medical research is undergoing its third paradigm shift of the last three centuries. Until the eighteenth and into the nineteenth centuries, research occurred informally in the course of treating a patient and observing outcomes—a physician learned from doing.[1] The clinical trial became the staple of medical research in the twentieth century.[2] The clinical trial remains important in the twenty-first century, but breakthroughs are increasingly coming from "informational" or "secondary" research,[3] that is, research that aggregates information about patients, including physical conditions, genetic information, treatments, responses, and outcomes. This research gives researchers a real-world snapshot at a population-wide level in a way that is not possible with traditional clinical trials. Data from clinical contexts are fed back into databases in a "continuous feedback loop" that iteratively helps improve clinical and health delivery outcomes.[4]

The new form of research is prominently foregrounded in both private and public initiatives, but requires vast quantities of information.[5] Numerous private payers, major health systems, and data intermediaries aggregate data which they use and sell to others.[6] The federal government has put into place its own programs. A 2001 Report put forth a vision for the National Health Information Infrastructure (NHII) which "would connect the multitude of participants in the health sector…and provide the means for

---

[1] Emily Largent et al., *Can Research and Care be Ethically Integrated?*, 41 HASTINGS CTR. REP. 37 (2011).

[2] Barbara J. Evans, *Much Ado About Property Ownership*, 25 HARV. J.L. & TECH. 69, 76 (2011).

[3] While my usage is standard, other scholars prefer different terms. *See, e.g.,* INSTITUTES OF MEDICINE, BEYOND THE HIPAA PRIVACY RULE 19 n.11 (2009) (hereinafter BEYOND). This Article is solely about secondary research based on data that is identifiable in some way, so that records about a single patient collected from multiple sources and at different points in time are linkable. *See,* Sharona Hoffman & Andy Podgurski, *Balancing Privacy Autonomy and Scientific Needs in EHR Research*, 65 SMU L. REV. 85, 130-31 (2012) (elimination of identifiable elements reduced data by 31 % and limited research). The Article does not concern research on biospecimens or clinical trials.

[4] INSTITUTES OF MEDICINE, INTEGRATING RESEARCH AND PRACTICE 13 (2014) (hereinafter INTEGRATING).

[5] *See* Federal Policy for the Protection of Human Subjects, 80 Fed. Reg. 53,933, 53938 ff. (2015) (hereinafter NPRM); Sharona Hoffman & Andy Podgurski, *Improving Health Care Outcomes Through Personal Comparison of Treatment Effectiveness Based on Electronic Health Records*, 39 J. L. MED. & ETHICS 425, 425 (2011); *see also* Lawrence O. Gostin, *Health Information Privacy*, 80 CORNELL L. REV. 451 (1995) (older but useful overview).

[6] Terry, *Big Data Proxies and Health Privacy Exceptionalism*, 24 HEALTH MATRIX 65 (2014).

managing the massive volumes of health data, information, and knowledge."[7] By mid-2014, as part of a *pilot* program, the Food and Drug Administration's (FDA) Sentinel post-market drug surveillance program could access the prescription records of nearly 180 million Americans, with over 48 million active records.[8] The Centers for Medicare and Medicaid (CMS) is in the process of implementing a program to digitize Medicare and Medicaid records through a payment incentive program for providers.[9] The programs together will collect data of 100 million Americans.[10] The Affordable Care Act's (ACA) still-developing Patient-Centered Outcomes Research Network (PCORNet) will access detailed data from 25 million patients.[11] Data in this program can range from basic clinical data, to genetic, demographic, and even behavioral information to study correlations and interactions.[12] Information could also be collected from healthy individuals to determine how they stay healthy, to help set both clinical and public health agendas.[13] The networks of private entities are even more sophisticated.[14]

Such mass data collection imposes burdens on individuals.[15] Apart from requiring them to understand and sign consents, or subjecting them to monitoring, there are privacy risks. "No security measures…can ever completely safeguard against…release…or inappropriate use."[16] Patients

---

[7] Don Detmer, *Building the National Health Information Infrastructure*, 3 BMC MEDICAL INFORMATICS AND DECISION MAKING 1 (2003) (citing NATIONAL COMMITTEE FOR VITAL AND HEALTH STATISTICS, INFORMATION FOR HEALTH: A STRATEGY FOR BUILDING THE NATIONAL HEALTH INFORMATION INFRASTRUCTURE (2001)).

[8] Sentinel Initiative Public Workshop, Engelberg Center for Health Reform at Brookings (Feb. 5, 2015), *available at* http://www.brookings.edu/~/media/events/2015/02/05-fda-sentinel-initiative-workshop/2015-sentinel-initiative-annual-meeting-slide-deck.pdf (double counting individuals who change plans.

[9] *See infra* note __.

[10] Centers for Medicare & Medicaid Services, *Medicare & Medicaid Statistical Supplement: 2012 Edition*, CMS, *available at* http://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/MedicareMedicaidStatSupp/2012.html.

[11] INTEGRATING, *supra* note 4, at 19.

[12] *See, e.g.,* Don Detmer, *Building the National Health Information Infrastructure*, 3 BMC MED. INFORMATICS & DECISION MAKING 1, 4 (2003).

[13] *See infra* notes 156-157.

[14] My description is, of course, a highly simplified one, that does not address several complications with each of these programs. Sentinel, for example, is better understood as public-private partnership, rather than a government run program per se. The various IOM reports to which I allude provide a good overview of these efforts.

[15] *See generally* sources cited *infra* note 18.

[16] PRECISION MEDICINE INITIATIVE (PMI) WORKING GROUP REPORT, THE PRECISION MEDICINE INITIATIVE COHORT PROGRAM, BUILDING A RESEARCH FOUNDATION FOR 21ST CENTURY MEDICINE

therefore face actual or feared employment or insurance discrimination, reputational loss, or identity theft.[17] Thus, those scholars writing on secondary research from an ethics-based viewpoint focus on the loss of privacy, and advocate limiting data collection.[18]

Pragmatic opponents respond by touting the benefits of information collection, which they say outweigh privacy harms.[19] This argument is simple, but overwhelmingly powerful. The benefits of secondary research promise to be legion. Agglomerating data has allowed researchers to identify genetic mutations that presage high risks of breast cancer or Alzheimer's,[20]

(Sep. 2015) (hereinafter PMI Report). The harms can range from reidentification to data breaches.

[17] Ryan Calo, *The Boundaries of Privacy Harm*, 86 IND. L.J. 1131 (2010); *see also* Konnoth, *An Expressive Theory of Privacy Intrusions*, 102 IOWA L. REV. (forthcoming 2017) (explaining that there are status harms).

[18] *See* Terry, *supra* note 6, at 99-100 (suggesting that health privacy exceptionalism requires more privacy protections); Nicolas Terry, *Protecting Patient Privacy*, 81 UMKC L. REV. 1, 5 (2012); Hoffman & Podgurski, *supra* note 3; Sharona Hoffman & Andy Podgurski, *In Sickness, Health, and Cyberspace: Protecting the Security of Electronic Private Health Information*, 48 B.C. L. REV. 331 (2007); Nicolas P. Terry & Leslie P. Francis, *Ensuring the Privacy and Confidentiality of Electronic Health Records*, 2007 U. ILL . L. REV. 681, 690 (2007); Terry, *Electronic health records: International, Structural and Legal Perspectives*, 12 J.L. & MED. 26 (2004); Peter D. Jacobson, *Medical Records and HIPAA*, 86 MINN. L. REV. 1497, 1497-99 (2002).

[19] Though they may endorse different weighting of the competing interests, some arguing for higher emphasis on privacy, others for research. *See, e.g.*, 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."); LAWRENCE GOSTIN, PUBLIC HEALTH: POWER DUTY, RESTRAINT 325 (2000); David Orentlicher, *Making Research a Requirement of Treatment: Why We Should Sometimes Let Doctors Pressure Patients to Participate in Research*, 35 HASTINGS CTR. REP. 20 (2005) (providing a balancing approach).

[20] Assn. for Molec. Pathology v. Myriad Genetics, 132 S. Ct. 1295, 2117 n. 4 (2013); *see* Anna Laakmann, *The New Genomic Semicommons*, 5 U.C. IRVINE L. REV. (forthcoming 2015). For a longer list that is screenable through, for example, PGD, *see What We Test For*, GENESIS GENETICS, http://genesisgenetics.org/pgd/what-we-test-for/; Fred Cate, *Protecting Privacy in Health Research*, 98 CAL. L. REV. 1765, 1780 (2002) (mutations that contraindicate the use of the popular blood thinner, warfarin); CANADIAN INSTITUTE FOR HEALTH INFORMATION & CANADA HEALTH INFOWAY, BETTER INFORMATION FOR IMPROVED HEALTH: A VISION FOR HEALTH SYSTEM USE OF DATA IN CANADA (2013), *available at* http://www.cihi.ca/cihi-ext-portal/pdf/internet/hsu_vision_report_en (sepsis in newborns); Simon et al., *Large Medical Databases, Population-Based Research, and Patient Confidentiality*, 157 AM. J. PSYCHIATRY 1731, 1734 (2000) (adverse effects of older antidepressants and sedative-hypnotic drugs);

changes to drug choice and administration,[21] and quality and cost control measures.[22] It promises to battle discrimination and stigma by revealing healthcare disparities and the commonness of certain conditions,[23] to help recruitment for clinical trials, and to enable research where trials are not possible.[24] These benefits are just the tip of the iceberg.[25] The ultimate goal is to create what policymakers call a "learning health system," where each clinical intervention will feed back into centrally accessible databases that in turn will set treatment standards, the results of which will again be fed back into the system.[26] Treatment will be optimized according to patient genetic and behavioral profiles, geographic costs and infrastructure limits, and staffing needs, among other variables. And of course, data will be used to craft policy.[27]

Given these benefits, the outcome of the debate between privacy and pragmatism is largely a foregone conclusion. The bargain has largely been struck, the deal made. Much scholarship simply takes for granted that data

---

John Bell, *The New Genetics: The New Genetics in Clinical Practice*, 316 BRIT. MED. J.618 (1998) (new understandings and definitions of disease based on physiological progression).

[21] Because of genetic and behavioral differences as individual variation in drug absorption, metabolism and elimination become clearer. Allen D. Roses, *Pharmacogenetics and Future Drug Development and Delivery*, 355 LANCET 1358 (2000); Zisis Kozlakidis et al., *Human Tissue Biobanks*, 8 RES. ETHICS 113 (2012) (noting that certain HIV positive individuals with the the HLA-B*5701 gene and Hepatitis C positive individuals with the IL28Beta gene should not receive certain medication).

[22] ALEX PENTLAND ET AL., BIG DATA AND HEALTH 31 (2013) (readmissions correlated with mental depression in Washington D.C. hospitals); *Id.* at 31 (brand name medication). L.O. Gostin, *Health Information: Reconciling Personal Privacy with the Public Good of Human Health*, 9 HEALTH CARE ANAL. 321, 322 (2001) (saving Britain's national health service $1.6 billion).

[23] *See generally* Simon et al., *supra* note 20.

[24] Tracy Stuardi et al., *Database Recruitment: A Solution to Poor Recruitment in Randomized Trials?*, 28 FAM. PRAC. 329 (2011) (discussing database recruitment); Walter F. Stewart et al., *Bridging The Inferential Gap: The Electronic Health Record And Clinical Evidence*, 26 HEALTH AFF. w181 (2007) (shortcomings with RCTs include that they are too selective and ignore comorbidities. Secondary research helps bridge the gap).

[25] E.M. Meslin & M.K. Cho, *Research Ethics in the Era of Personalized Medicine: Updating Science's Contract with Society*, 13 PUB. HEALTH GENOMICS 378 (2010) (explaining how limitations in secondary research structure has limited various discoveries). The IOM's Roundtable on Value & Science-Driven Health Care has set a goal that, by the year 2020, 90 percent of clinical decisions will be supported by accurate, timely, and up-to-date clinical information based on secondary research. INSTITUTES OF MEDICINE, REDESIGNING THE CLINICAL EFFECTIVENESS RESEARCH PARADIGM xv (2010).

[26] INSTITUTES OF MEDICINE, BEST CARE AT LOWER COST: THE PATH TO CONTINUOUSLY LEARNING HEALTH CARE IN AMERICA (2012).

[27] *See generally id.*

collection will occur. Policymakers recognize the value of privacy only where it does not interfere with these important goals.

The question then, is not *whether* we will incur privacy risk, but *how* we will distribute risks, benefits and burdens across society. Outside the information context, questions of distributional justice, legitimacy, and efficiency are of central importance in health debates.[28] But in the battle between privacy and pragmatism, these other values have largely been ignored.[29] Information collection has generally taken the path of least resistance, occurring where there is already stakeholder support (proponents of cancer research) or large centralized systems (Medicare/Medicaid data).[30] There is little coordination among programs such as PMI, PCORNet, CMS Meaningful Use, and FDA Sentinel or private counterparts.[31] Capacity building and scholarship focus on only technological expertise, improvements,

---

[28] *See, e.g.*, NORMAN DANIELS, JUST HEALTH (2008).

[29] The exception here is Lior Strahilevitz, Symposium, *Toward a Positive Theory of Privacy Law*, 126 HARV. L. REV. 2010 (2013). However, Strahilevitz focuses on the very different question of who wins or loses when information is regulated, rather than the distribution of privacy risks among the population.

[30] PCORNet, similarly, consists of twenty-nine large private clinical networks or health systems and patient networks, which have already built 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. *See* PCORNET, PCORNET GOVERNANCE (2013) *available at* http://www.pcornet.org/patient-powered-research-networks/community-and-patient-partnered-centers-of-excellence-phase-ii/; PMI Report, *supra* note 16, at 28, 30

[31] PCORNet includes individuals from other federal agencies, including the Office of the National Coordinator for Health Information on its Board. *Id.* But other entities do not. FDA Sentinel for example, includes no individuals from other agencies even though private entities are well represented. FDA, MINI-SENTINEL PRINCIPLES AND POLICIES 8 (Nov. 2014). PMI similarly calls for "cross agency coordination" but final authority rests with yet another component of HHS, the NIH Director. PMI Report, *supra* note 16, at 90.

and incentives,[32] and improving the outcomes of specific programs, rather than on ethics.[33]

This ad hoc, technology-focused approach distributes privacy risks in ways that are unjust and illegitimate. CMS Medicare/Medicaid data, for example, which contains the claims of the elderly and poor, is, by law, widely "made available" for research.[34] Under programs such as PCORNet and other

---

[32] OFFICE OF THE NATIONAL COORDINATOR FOR HEALTH INFORMATION TECHNOLOGY, REPORT TO CONGRESS, UPDATE ON THE ADOPTION OF HEALTH INFORMATION TECHNOLOGY AND RELATED EFFORTS TO FACILITATE THE ELECTRONIC USE AND EXCHANGE OF HEALTH INFORMATION 18-19 (2014); DEP'T OF HEALTH AND HUMAN SVCS., EHR PAYMENT INCENTIVES FOR PROVIDERS INELIGIBLE FOR PAYMENT INCENTIVES AND OTHER FUNDING STUDY (2013). For scholarship, see Nicholson Price, *Blackbox Medicine*, 28 HARV. J.L. & TECH. 419, 424 (2014) (intellectual property). Others focus on improving technology and data quality include Hoffman & Podgurski, *supra* note 5; Nicolas P. Terry, *Meaningful Adoption: What We Know or Think We Know About the Financing, Effectiveness, Quality, and Safety of Electronic Medical Records*, 34 J. LEGAL MED. 7 (2013); Sharona Hoffman & Andy Podgurski, *Finding a Cure, The case for Regulation and Oversight of EHR Systems*, 22 HARV. J.L. & TECH. 1 (2008).

[33] Sharona Hoffman & Andy Podgurski, *E-Health Hazards: Provider Liability and Electronic Health Records*, 24 BERKELEY TECH. L.J. 1524 (2009) (liability problems that arise with HER systems); Evans, Authority of the Food and Drug Administration to Require Data Access and Control Use Rights in the Sentinel Data Network, 65 Food & Drug L.J. 65 (2010); Evans, *Institutional Competence to Balance Privacy and Competing Values: The Forgotten Third Prong of HIPAA Preemption Analysis*, 46 U.C. DAVIS L. REV. 1175, 1192 (2013) (HIPAA sets a privacy law ceiling that its regulations ignore); Hoffman, *Citizen Science: The Law and Ethics of Public Access to Medical Big Data*, BERKELEY TECH. L.J. (forthcoming 2015) (problems created with public acess to big data); Evans, *Congress' New Infrastructural Model of Medical Privacy*, 84 NOTRE DAME L. REV. 585 (2009) (problems and solutions involving Sentinel system in balancing confidentiality versus public health); Evans, *The Ethics of Postmarketing Observational Studies of Drug Safety Under Section 505(o)(3) of the Food, Drug, and Cosmetic Act*, 38 AM. J.L. MED. 577 (2012); Barbara J. Evans, Waiving Your Privacy Goodbye: Privacy Waivers and the HITECH Act's Regulated Price for Sale of Health Data to Researchers, University of Houston/Health Law & Policy Institute Working Paper No. 2010-A-22, http://ssrn.com/abstract=1660582 (explaining the privacy problems the HIPAA's cost-based fee creates). The few exceptions in the literature that consider how best to create a governance model focus purely on who should control the data, without considering questions of distributional justice. *See* Evans, *supra* note 2 (discussing question of data ownership);

[34] 42 U.S.C. 1320e(d)(3)(A) (Affordable Care Act). The federal Healthcare Cost Report Information Systems began providing publicly available information for Hospitals in 1996, JHAs and Renal in 1994, and Hospice from 2000 onward. FQHC/RHCs in 2010. *See, e.g.*, HCRIS, Frequently Asked Questions, https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/Cost-Reports/DOCS/HCRIS-FAQ.pdf (last accessed Nov. 6, 2015); CMS, Cost Reports, https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/Cost-Reports/?redirect=/costreports/02_hospitalcostreport.asp (last accessed Nov. 6, 2015). Making these data available subject them to breaches more readily than private

exchanges, private entities make available deidentified data of their clients without their consent. Only those individuals who can pay out of pocket, or otherwise have meaningful choice with respect to health delivery can opt out of the more intrusive forms of data sharing.[35]

This Article shifts the focus from *whether* we should allow privacy risks, to *how* privacy risks can justly, legitimately, and efficiently be distributed across society. Instead of pragmatic demand side questions—what research questions do we want answered and how can we most easily get the data to do so, it asks ethical supply side questions—how and from whom do we justly get the data for the research. The answer implicates other values besides privacy, including substantive justice, procedural legitimacy, and efficiency. All these values impose mutual constraints on each other—majoritarianism may be the most *legitimate* approach, but may be inequitable and inefficient; but imposing the most *equitable* system on an unwilling populace may be illegitimate. Information collection processes must therefore be conceptualized as part of a single system that takes into account all these values. To do so, we must look elsewhere for inspiration.

The collection of money in the income tax system offers the most useful model for the new health information collection system. As a descriptive manner, like the new health information collection system, taxation involves collecting resources routinely and universally from the population at large for the public good. As a normative matter, bioethicists and those who justify taxation rely on similar principles to justify this collection. An individual must: repay society for the benefits she gains from it (the benefit principle), contribute to aid those less well off than her (the redistribution principle), and pay in order to show solidarity with fellow citizens (the solidarity principle).

The analogy serves three purposes. The first is conceptual. The ethical myopia that has plagued information collection is the result of near-sighted focus on the technical requirements of specific programs instead of on

---

proprietary data. Further, CMS is not fully complying with breach notification requirements and does not offer remedies to the Medicare patients when there are breaches. Pamela Dolan, *Investigation Faults Handling of Medicare Patient Data Breaches*, AMEDNEWS.COM, Oct. 29, 2012 (summarizing results of HHS OFFICE OF INSPECTOR GENERAL, CMS REPONSE TO BREACHES AND IDENTITY THEFT (2012)) *available at* http://www.amednews.com/article/20121029/business/310299965/6/; *see also* GOV'T ACCOUNTABILITY OFC., REPORT: CENTERS FOR MEDICARE AND MEDICAID SERVICES NEEDS TO PURSUE A SOLUTION FOR REMOVING SOCIAL SECURITY NUMBERS FROM CARDS (2013).

[35] This choice is not necessarily tied to wealth. It may arise from greater competition in their respective health exchange, the size of a health-care subsidy under the ACA, employer-provided options, and various other criteria.

the system as whole. The analogy helps us conceptualize all information collection programs as components of a single system—just as taxation of capital gains, rent, or salaries are all part of the income system, so too are the information streams from CMS, FDA, and PCORNet the building blocks of a single NHII. Such clarification would hopefully result in policy changes that consolidate fragmented data collection channels. This vision of a broader system formed the basis of recent comments to HHS, where I argued against policies that "silo[]" away certain kinds of information from the "broader system."[36] In future efforts, HHS should consider consolidation of PMI, PCORNet, and Sentinel programs under the oversight of a single entity.

Framing the collection enterprise as a single system lays the groundwork for the second goal of this Article—developing systemic principles of substantive and procedural justice that will undergird the system. First, as a matter of substantive justice, equity demands that individuals with lower welfare suffer fewer tax burdens and enjoy greater benefits. Similarly, I argue that the distribution of privacy risk in the health information system should track individual welfare. Where possible, and all else being equal, individuals with lower welfare because of their income, age, family size, health, etc. should be subject to lesser privacy risk. This principle would require immediate policy changes—for example, shifting the burdens of information collection and privacy risks away from the Medicaid/Medicare populations where they are currently concentrated;[37] and redirecting PCORNet and PMI grants towards data sources that have been unfairly underexploited, such as wearable technology.[38] As the broader NHII coalesces, the principle would guide risk distribution in a more systematic manner.

As a procedural matter, the tax analogy reinforces changes we are already seeing in consent mechanisms for information collection. Tax collection does not demand individual informed consent. Rather, tax

---

[36] Craig Konnoth, Comment on the Confidentiality of Substance Use Disorder Patient Records, 81 Fed. Reg. 6987, Docket ID number SAMHSA-4162-20 at 2 (2016) (arguing for alignment between these protections and HIPAA rules).

[37] I do not mean to suggest that CMS's incentive programs will not be relevant for such initiatives; indeed, those programs help drive technology transformation to achieve several of the goals I describe here.

[38] PMI is considering this question. *See* PMI Report, *supra* note 16, at 14 ff. Other contexts abound. For example, the Department of Labor may be called upon to write data collection standards for state all payers claims databases. These databases were recently found preempted by ERISA in Gobeille v. Liberty Mutual Insurance Company, No. 14-181 (U.S., Dec 2, 2015).

principles require that society as a whole debate tax policy and elect representatives who enforce the decision of the community on all of society. Similarly, we should collect health information using this democratic consent model, where society as a whole settles on data collection projects to which all individuals are subject. This would require changes to existing informed consent requirements for information collection under the Health Insurance Portability and Accountability Act of 1996 (HIPAA),[39] and the proposed reform to the rules governing all research in all federally funded institutions.[40]

The third goal of the analogy is to engage scholars, policymakers, and eventually, the general public, with the ethics of health information collection in general, and perhaps, specifically, with the principles I outline. The new system needs investment from everyone. But so far, engagement has been difficult. Most consider health information collection policy to be esoteric and technocratic.[41] The analogy helps provide a useful initial frame to show that there are important and understandable ethical principles at stake. Everyone understands the rudiments of taxation; everyone cares about how the burdens and benefits of taxes are distributed, and nearly everyone subscribes to its tenets, even if some do so grudgingly. Only through such engagement will individuals reliably disclose medical information,[42] and even

---

[39] 110 Stat. 1936. The relevant regulations are at 45 C.F.R. Part 160 and Subparts A, C & E of Part 164.

[40]

[41] A. Hobbs et al., *The Privacy-Reciprocity Connection in Biobanking*, 15 PUB. HEALTH GENOMICS 272, 278 (2012) argues based on UK biobank-related interviews that individuals want information regarding how their data helped in exchange for the loss of privacy. As some noted, "we are giving a lot more information than they are giving us." *Id.* at 280. In response, the administrators have proposed listing uses of research on the website to allow for more public input and to increase collection; *See also* Helen Busby & Paul Martin, *Biobanks, National Identity & Imagined Communities*, 15 SCI. AS CULTURE 237, 243 (2006) (health information policy discussions tend not to be public).

[42] Individuals have been found to avoid health information collection by declining to go to the doctor or engaging in self-medication. *See* Deborah Peel, *The Case for Informed Consent,* CONSUMER WATCH DOG 5 (Aug. 2010), http://www.consumerwatchdog.org/sites/default/files/resources/peel.pdf. Similarly, Carol Diamond et al have argued that doctors and patients fail to give information because they are not included in the overall system—they provide information and never see the benefits or hear what has happened to it. They infer that they feel used and treated like means—in other words, disrespected. IRBs make decisions as to individuals' information without their knowledge, the entire process is a blackbox in which the individual cannot participate. Carol C. Diamond et al., *Collecting And Sharing Data For Population Health: A New Paradigm*, 28 HEALTH AFF. 454, 458 (2009).

take steps to make sure that their medical records are accurate to support research.[43] As time goes on, we can move beyond reliance on the tax analogy, to discuss health information collection in its own right.

Two caveats are in order. First, this Article takes no view on *whether* we should collect information, *if* collection continues, the analogy provides an ethical framework within which to reorganize privacy risk.[44] Second, the analogy is not, of course without its weaknesses. Taxation has its critics and its problems, to be sure. And health information collection may find closer analogies elsewhere than in revenue taxation. But as other authors have shown, and as I hope to show here, analogizing information resources to money or tangible property helps garner valuable insights.[45] Unlike other framing analogies, this one conceptually reframes the collection enterprise as part of a single system and provides the basis for cross-cutting principles of substantive and procedural justice. Further, although there is disagreement on specific points, the tax principles on which I base my analysis have been politically vetted and are generally, if grudgingly, accepted by individuals with different conceptions of justice.

Part I explains why older models of health information collection provide weak ethical guides for new collection approaches. Part II presents the basis for the taxation analogy, showing how health information collection and taxation track each other both descriptively and normatively. Part III presents the ethical framework that springs from this analogy.

I. THE NEW MODEL OF HEALTH COLLECTION

---

[43] Sharona Hoffman, *Medical Big Data*, 21 CONN. INS. L. J. 289, 307 (2014).

[44] The analogical approach is a pragmatic approach to developing ethical principles to guide health information collection. Unlike foundational and abstract theories of justice, such as that of John Rawls or Robert Nozick, NORMAN DANIELS, JUST HEALTH (2008) for example deduces principles similar to mine by applying a Rawlsian framework. Daniels argues that a healthcare system should be (1) equitable, (2) accountable, and (3) efficient. Id. at 248-254. In this respect, my argument (arrived at independently of Daniels work) tracks Daniels. However, as Daniels explains, to accept his argument, one must explicitly hew to a Rawlsian philosophy. Id. at 103. By contrast, my framework contains no such limitation. Taxation commands sufficient consensus and has been vetted over centuries by scholarly and public debate. All (or nearly all) of us recognize its necessity, if begrudgingly. Certain key similarities between taxation and health information collection mean that principles of justice that are applicable in the tax context play some role in the health information context, whatever someone's underlying ideological commitments are.

[45] *See infra* note ___.

Modern health information collection departs sharply from traditional models of health information in key areas, both descriptive and normative. Accordingly, traditional health information cannot provide a model for the new health information collection enterprise.

A. Descriptive Characteristics

Traditional health information collection occurs in two, non-routine, specialized contexts—randomized controlled clinical trials (RCT) or for public health surveillance. With certain exceptions such as vaccine reporting, both forms of collection focus on a specific group of individuals. The burdens from collection may often be heavy. As I have described elsewhere, collection of information for public health singles out the group concerned, and, often stigmatizes it.[46]

Next, the purpose of the surveillance is predefined. The surveillance must be directed at a specific goal, such as examining a particular hypothesis, or addressing a particular, existing or imminent, public health problem. A court or Institutional Review Board (IRB) evaluates whether the purpose is sufficiently definite.[47] The information collected from the individual has a direct relationship with that specific goal, and cannot be repurposed. Thus, information collected about tuberculosis may be used only to quarantine an individual reported as contagious to prevent the spread of tuberculosis.

The new system differs in three key ways. First, routine clinical care is the backbone of this system. Medicaid, Medicare, and prescription data, for example, is collected as part of the clinical encounter. That data is agglomerated into a central holding system, or is accessed remotely, as I describe further in the next Parts. The data is used to develop correlations between individuals' characteristics and the ways in which they experience illness and respond to treatment. That information is then fed into clinical decision support (CDS) systems, which guide doctors on how best to treat

---

[46] Craig Konnoth, The Burdens of Public Health Surveillance (unpublished manuscript, 2015).

[47] *See* 45 C.F.R. 46.111(a)(1) (1998) and 45 C.F.R. 46.113 (1998); Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data,1995 O.J. (L 281) 31, art. 6(1)(b), *available at* http:// www.europa.eu.int/eur-lex/en/search/search_oj.html (requiring "specific" and "explicit" purposes, after achieving which the collection and use must be stopped).

patients with that profile.[48] The information from that next encounter is then fed back into the system, in an iterative loop, continuously refining its insights.

Accordingly, nearly everyone contributes to the collection. No group is singled out for, or stigmatized by, collection. Further, the collection need not be directed towards a specific purpose. For example, unlike traditional RCTs, observational research may proceed without a preexisting hypothesis that is then tested for. Rather, computational methods can be used to identify correlations that had not previously been considered.[49] Even if a hypothesis is pre-identified, longitudinal information from a patient's past may prove important. Without ensuring that records are collected ex ante without any particular purpose in mind, such information may be unavailable.

B. Normative Characteristics

Next, the normative justifications of the new collection are more expansive than traditional collection in two ways. First, they provide an account of individual obligations to society. Second, they implicate unexamined questions of social, political, and economic justice.

RCT or public health surveillance consider only a particular purpose at a particular point in time. They therefore do not consider the interests of individuals and society as interactive and mutually-shaping. The Common Rule and the Health Insurance Portability and Accountability Act (HIPAA) require IRBs to balance the level of the risk versus the degree to which the waiver and information are vital to the research.[50] Similarly, in the limited case law on public health surveillance, courts balance the right of privacy against other rights.[51] Most scholars endorse this balancing approach even

---

[48] *See* INSTITUTES OF MEDICINE, *supra* note 26**Error! Bookmark not defined.** and related discussion.

[49] W. Lipworth, *Reconceptualizing Tissue Banking Consent*, 36 INTERNAL MED. J. 124 (2006).

[50] 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)(1) (1998) and 45 C.F.R. 46.111(a)(2) (1998); Standards for Privacy of Individually Identifiable Health Information, 65 Fed. Reg. 82,464 (Dec. 28, 2000) (HIPAA represents a balance).

[51] JAMES HODGE ET AL., LEGAL ISSUES CONCERNING IDENTIFIABLE HEALTH DATA SHARING BETWEEN STATE/LOCAL PUBLIC HEALTH AUTHORITIES AND TRIBAL EPIDEMIOLOGY CENTERS IN SELECTED U.S. JURISDICTIONS (2011); Margaret Hoppin, *Overly Intimate Surveillance: Why Emergent Public Health Programs Deserve Strict Scrutiny under the Fourteenth Amendment*, 87 N.Y.U. L. REV. 1950 (2012).

with respect to modern health information collection projects.[52] The balancing analysis treats individual and social interests as independent, even opposed to each other.[53] Only one may lay final claim to the information; either individual rights can be respected, or communal goals.

The new collection relies on more expansive accounts of individual-social interaction. The new collection spans vast populations, and involves collection and re-collection projects in iterative loops, rather than a single trial or public health event. Over these larger expanses of time and space, the interests of individual and society are mutually constituted and developed through individual-social interaction: the individual provides information to aid society, which then creates products to help the individual. The information therefore belongs to not one entity, but rather, is "in a gray area between a public and private good;"[54] it flows through a system that belongs simultaneously to both citizens and society and which represents their interlocking fates.[55]

Because of this mutuality, the new collection must and often does rely on justifications that go beyond frozen-in-time balancing calculi. In the health information field, admittedly, policy makers and many scholars continue to rely largely on balancing analyses.[56] But in the broader literature on medical research, ethicists have long argued that individuals must contribute to scientific research. The reasons—which I address only briefly here and expand upon in the following Part—roughly fall into three

---

[52] Though they may endorse different weighting of the competing interests, some arguing for higher emphasis on privacy, others for research. *See, e.g.*, 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."); LAWRENCE GOSTIN, PUBLIC HEALTH: POWER DUTY, RESTRAINT 325 (2000); David Orentlicher, *Making Research a Requirement of Treatment: Why We Should Sometimes Let Doctors Pressure Patients to Participate in Research*, 35 HASTINGS CTR. REP. 20 (2005) (providing a balancing approach).

[53] Bruce Jennings, *On Authority and Justification in Public Health*, 55 FLA. L. REV. 1247 (2003).

[54] INSTITUTES OF MEDICINE , CLINICAL DATA AS THE BASIC STAPLE FOR A LEARNING HEALTH SYSTEM 253 (2010) (hereinafter, BASIC STAPLE).

[55] Nonetheless, because of the balancing approach, the extremes rule the day. *See, e.g., id.* at 53 ("Under one scenario, health information could become a true public good as something that is truly nonproprietary. Under another scenario, clinical information could become a private good as something that is used differentially, for comparative advantage that benefits some, but not all.").

[56] *See supra* note 52.

categories. First, because individuals benefit from the fruits of mass collection, fairness demands that they contribute back to society. Second, individuals have a duty, analogous to a duty to rescue, to aid those in worse straits than them.[57] Third, individuals contribute as an act of solidarity; their contributions represent the common biological heritage we all share. Underlying each rationale is a normative vision of individual-social interaction.

Finally, because traditional health information collection focuses on a single moment of collection aimed at specific purposes, it largely ignores larger questions of social, economic, and political justice. But the new health information collection collects vast quantities of information from large groups of individuals, for broad, sometimes unforeseen or unforeseeable purposes. Thus, the collection results in *collective* resources on one hand and *collective* benefits on the other—a wholesale rather than a retail approach.

We therefore cannot think of the new collection as a series of ad hoc decisions with a single burden balanced against a specific and related benefit—there is no one-to-one correspondence. Rather, we must generate an optimal mix of burdens and benefits, that most conforms to our set of preferences across society as a whole. This mix may involve less overall privacy in some circumstances and more in others, fewer health benefits in some contexts and more in others. Modern health collection practices should be understood as forming a single system, rather than a series of ad hoc practices, that generates the appropriate basket of burdens and benefits. These burdens and benefits must be legitimately and justly distributed. This implicates questions far beyond the purview of the old balancing approach.

II. THE CASE FOR A TAXATION ANALOGY

So far I have identified only those characteristics that differentiate the new health information system from the old. Other descriptive and normative characteristics also define the new enterprise. These

---

[57] Some bioethicists argue that individuals have a general duty to provide tissue and information for research but the debate has focused on *whether* there is a duty. As a result, while writers have provided numerous reasons for research, these reasons are not systematized or organized. John Harris, *Scientific Research is a Medical Duty*, 31 J. MED. ETHICS 242 (2005); Iain Brassington, *John Harris' Argument for a Duty to Research.*, 21 BIOETHICS 160 (2007); Sarah Chan & John Harris, *Free Riders and Pious Sons - Why Science Research Remains Obligatory*, 23 BIOETHICS 161 (2009); Iain Brassington, *Defending the Duty to Research,* 25 BIOETHICS 21 (2011)

characteristics together find a close analogy in the defining aspects of the collection of money in the income taxation system.[58]

To be sure, money and information are resources with different properties in themselves. Money can be translated into a larger variety of goods and services ranging from bridges to buildings, than health information, which can only be used to improve healthcare delivery. Money is also exhaustible—once it is collected, the individual from whom it is collected no longer has access to it. Information is non-exhaustible in this manner. Unlike in the case of revenue, where a tax once paid melds into the pool of revenue paid in by other individuals, health information in some ways may remain connected to the individual who gave it even if the information is ultimately deidentified, more so if it is not. In the revenue context, one person's contribution does not differ in any meaningful way from another person's. In the health context, while the same kinds of information are collected, the information itself differs from other individuals' (indeed, that is the reason we collect it).

All the same, a rich vein of literature has used frameworks that treat information like money or other kinds of property to yield insights or prescribe efficient or ethical frameworks. Many suggest that information is or should be used as consideration to "pay" for goods and services, or that information should be bought and sold on the open market.[59] Others prescribe information commons or communal ownership, explicitly using models developed for tangible property.[60] My own model goes one step further—if information can be used as a form of payment, it can also be "taxed." Although money and information are different resources *in themselves*, the systems they belong to can delineate *relationships* between individual and society or among individuals that are similar in important ways. The systems impose comparative burdens on individuals, involve near-

---

[58] Some of the characteristics of the health information system that I describe are still in their nascent stage. Next, while some may choose to analogize to other forms of taxation, such as consumption taxation, such forms of taxation do not share the same ethical commitment to, say, redistribution, as income taxation, a commitment that is reflected in the bioethics literature.

[59] Mark A. Hall, Property, Privacy, and the Pursuit of Interconnected Electronic Medical Records, 95 Iowa L. Rev. 631 (2010) (arguing for a markets based approach where individuals can buy and sell information); Hall provides numerous cites to the large literature on the general analogies between money/property and information.

[60] Kathryn Strandburg, [forthcoming book] (creating an information commons approach based on a commons approach created for property); Rodwin, The case for public ownership of patient data, 302 JAMA 86 (2009).

universal collective pooling of resources, for the common good, and justify this collection on similar normative grounds.[61] This, in turn, leads to comparable ethical duties.

A. The Descriptive Case

1. Burdens

Collecting information and collecting money impose burdens on individuals. Taking money from individuals imposes on them the risk that they will be unable to pursue other goals such as purchasing objects or saving money. They may also experience psychological burdens that comes with the risk or from having to give up an endowed good.[62]

Although information is a different kind of resource than money, collecting information also subjects individuals to risks that their information may fall in the wrong hands and be used either to harm them or to produce goods of which they disapprove.[63] There is also the psychological burden of losing control over information. Individuals may also have to expend time and resources in order to allow a provider to collect their data.[64]

The burdens imposed by taxation and information collection are comparable in at least three ways—the manner in which they are distributed, their intensity, and their differentiation.

First, burdens are universal and routine. Part I made the case that that the new health information collection system increasingly operates in

---

[61] In building this analogy, of course, I only hit on major themes in the taxation literature—more specific points have either not been considered in the bioethics literature (and are addressed further in Part III) or are too complicated for this preliminary treatment.

[62] *See* Craig Konnoth, *Revoking Rights*, 66 HASTINGS LJ. 1365 (2015).

[63] Some of the problems of which I explore elsewhere. Konnoth, *supra* note 46; *see also See also* David Korn, *Medical Information Privacy and the Conduct of Biomedical Research*, 75 ACADEMIC MED. 963 (2000) (explaining this as pragmatic versus ideological concerns regarding information sharing). Scientists have attempted to quantify this risk. *See, e.g.*, Bradley Malin & Latanya Sweeney, *How (Not) to Protect Genomic Data Privacy in a Distributed Network: Using Trail Re-Identification to Evaluate And Design Anonymity Protection Systems*, 37 J. BIOMEDICAL INFORMATICS 179 (2004). However, there is no uniform approach to calculating risk. In its "Administrative safeguards" section, the Security Rule requires covered entities to conduct an accurate and thorough assessment of the potential risks and vulnerabilities to the confidentiality, integrity, and availability of electronic protected health information: No further details are provided concerning how the complex task of risk analysis should be accomplished 45 C.F.R. 164.308(a) (1) (ii) (A). Various antidiscrimination laws have been put into place to prevent this kind of

[64] Because this Article is restricted to *informational* research, I assume that none of the information collection involves risky medical procedures.

routine medical contexts. The ambition of the system is to collect information as universally as is practicable, though gaps remain when individuals do not see medical providers. Gaps in collection lead to problems with statistical bias that undermine research conclusions.[65] There are proposals in place to close these "loopholes."[66] Income tax is similarly collected annually and routinely rather than after a specific event—death for example, as in the case of estate taxes, or after a purchase, as with sales taxes. Unlike other kinds of taxes, like property taxes, we think of income taxes as universally applicable, albeit with certain exceptions or loopholes.

Second, the intensity of the burdens is comparable in important ways. Because money and information are different kinds of resources, it is hard to make a direct comparison. But both information and revenue collection differ from comparable burdens in similar ways.

Through income taxation, the state claims some portion of the goods citizens produce through their labor.[67] But the state exploits citizens' productivity in other ways that are considered more burdensome. Conscription for example, is generally regarded as a heavier burden on individuals than taxation. Next, information collection also involves the collection of resources owned by or otherwise connected to the citizen.[68] But other health research imposes heavier burdens on individuals.[69] Interventional research may be carried out through RCTs and may involve the ingestion or injection of substances, even surgery.[70] The burdens these risks entail are far greater than those of informational research which, at

---

[65] 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 biases statistical studies. This prevents scientists from drawing robust conclusions regarding disease profiles, and the effects of various behavioral, environmental, and other characteristics on health. Rodwin, *supra* note 210, at 610. BEYOND, *supra* note 4, at 209-213 provides a useful overview.

[66] *See, e.g.*, Craig Konnoth, Integrating Employer Wellness Programs, Opening Plenary Panel, Health Privacy Summit, Georgetown Law Center (June 3, 2015).

[67] *See, e.g.,* Daniel Markovits, *Luck Egalitarianism and Political Solidarity*, 9 THEORETICAL INQUIRIES LAW 271 (2007).

[68] Although the literature is unclear, I argue elsewhere that health information is best understood as owned by its subject. *See* Craig Konnoth, Health Information Ownership (on file with author).

[69] *See, e.g.*, NPRM, *supra* note 5; R. Faden et al., *Ethics and Informed Consent for Comparative Effectiveness Research with Prospective Electronic Clinical Data*, 51 MED. CARE S53 (2013).

[70] Wolf & Buckwalter, *Randomized Surgical Trials and "Sham" Surgery*, 26 IOWA ORTHOP. J. 107, 107 (2006).

most, includes information collected through surveys and minimally intrusive swabs.[71]

Conscription and RCTs are greater burdens than taxation or informational research, respectively, for similar reasons. Income taxation or information collection imposes incidental burdens: one would continue to earn money or go to the doctor whether or not the tax or information was collected. But with conscription and RCTs, the individual's action completely serves the goals of another entity, and none of their own (apart, perhaps, from altruistic satisfaction). Further, income taxation and information collection intrude far less upon bodily integrity than conscription or RCTs. The extent to which bodily intrusion is or is not involved presents an important ethical distinction.[72] Indeed, RCTs have frequently been compared to conscription in the bioethics literature.[73]

Third, both our income and information collection systems are extraordinarily sensitive to the source of the resource, calibrating collection to minimize the burden as appropriate. The income tax system differentiates between various kinds of income. Social security income, for example, remains untaxed.[74] Income earned through capital gains is taxed in different ways and at different rates than salaried income.[75] There are several other categories. There is sometimes overlap and controversy as to which category a particular kind of income falls into.[76] But the taxation is shaped by norms regarding how each kind of income should be taxed.

---

[71] *Cf.* Maryland v. King, 133 S. Ct. 1958 (2013).

[72] "The present question is not whether we think" about the body differently. "We do." Alan Wertheimer, *(Why) Should we Require Consent to Participation in Research*, 1 J.L. & BIOSCIENCES 137, 157 (2014); CHARLES FRIED, MEDICAL EXPERIMENTATION: PERSONAL INTEGRITY AND SOCIAL POLICY (1974). ("The human person identifies himself with his body; he knows that he IS his body, that his knowledge of and relation to the whole of the outside world depends on his body and its capacities, and that his ability to formulate and carry out his life plan depends also on his body and its capacities."); CECILE FABRE, WHOSE BODY IS IT ANYWAY? (2006). ("'the objection from bodily integrity derives much of its force from the view that in violating people's bodily integrity, one is interfering with their life to an unacceptable extent.").

[73] *See, e.g.*, Soren Holm et al., *Conscription to Biobank Research?, in* ETHICS OF RESEARCH BIOBANKING 255, 258 (Jan Helge Solbakk et al. eds., 2009).

[74] 20 C.F.R. § 404.

[75] 26 C.F.R 1.

[76] CCH Social Security Reporter. ¶ 17,729B CRS Report for Congress: Social Security: Calculation and History of Taxing Benefits. (July 18, 2006); *see also* Jerry W. Markham, *Privatizing Social Security*, 38 SAN DIEGO L. REV. 747 (2001)

As I explain in detail elsewhere,[77] health information is often differentiated and collected depending upon various social norms. Anonymized clinical information is regularly collected already. Personal identifiers, and other information with potential medical implications are not collected, though this approach may soon change.[78] Collecting other kinds of information such as food purchase or consumption that are produced in non-clinical contexts would be considered highly inappropriate even though they are relevant to our health.

2. Benefits

Taxation raises revenue for the public good. Public goods are usually in-kind—the pooled income is transformed into services consumed by the community. Like pooled money, pooled information is transformed through appropriate means into useful public goods in the form of medical breakthroughs.[79]

The goods of secondary research are public both as a common sense as well as a technical matter. As a common sense matter, medical breakthroughs, and cost and quality control measures benefit society as a whole. They are therefore public goods. Under the economic definition of public good, goods are public when they are non-rivalrous and non-excludable to some extent.[80] The classic public good is national defense. National defense is non-rivalrous because one individual's benefiting from it does not decrease the benefits others derive from it. It is non-excludable because free-riders cannot be excluded from the benefit.[81]

---

[77] Symposium, Beyond IRBs: Designing Ethical Review Processes for Big Data Research, *Classification Standards for Health Information: Ethical and Practical Approaches*, 72 WASH. & LEE L. REV. ONLINE 395 (2016); PMI Report, *supra* note 16, at 47-48.

[78] FAQs, Terms and Recommendations on Informed Consent and Research Use of Biospecimens, July 20, 2011, http://www.hhs.gov/ohrp/sachrp/commsec/attachmentdfaq's termsandrecommendations.pdf.pdf

[79] *See supra* note 49 and accompanying text.

[80] David Blumenthal, *Characteristics of a Public Good and How They are Applied to Healthcare Data, in* BASIC STAPLE, *supra* note 210. Indeed, many seeking to impose a moral obligation to participate in public health research (clinical or observational) rely on the notion that health research is a public good for which everyone must support. Stuart Rennie, *Viewing Research Participation as a Moral Obligation*, 41 HASTINGS CTR. REP. 40, 42 (2011).

[81] The economic problem with public goods is that without government intervention, there is no incentive to produce them. If free riders cannot be excluded, and if the use of the good

Former National Coordinator for Health Information Technology, David Blumenthal, among others, has argued that the health benefits that come from improved health research tools constitute public goods. As Blumenthal acknowledges, intellectual property law renders some knowledge gained from biomedical research rivalrous and excludable up to a point.[82] But as a condition of IP protection, some elements of the scientific research become public knowledge and can be used for further research.[83] And often, general insights into correlations between certain physiological and behavioral characteristics and medical conditions become publicly available information. This knowledge may make healthcare cheaper, or more effective for everyone.[84]

3. Government Intervention

The taxation system has evolved over time. In the past, the collection of tax revenue was contracted out to private entities.[85] As time went on, governments took the lead on tax collection, with some assistance from private entities such as employers and investment entities, who withhold taxes as required by law.[86]

A similar move is in the works in the health collection sphere. To be sure, in the United States, the government does not hold as much of a monopoly over health information collection as it does over tax revenue collection. There are multiple causes for this, such as the lack of a single payer system as in other nations. Even those who recognize the need for a centralized repository or access mechanism promote private intermediaries, rather than the government.[87]

---

does not prevent others from using it, it is hard to charge for the good. Not all public goods are equally as non-rivalrous or non-excludable, but tend to exhibit these characteristics to some degree or another. *Id.*

[82] *Id.*

[83] *Id.*

[84] Note that I argue here that only the knowledge produced is a public good; the healthcare activities that apply the knowledge to actual patients are not public goods under the technical economic definition of the term.

[85] *See* Nicholas R. Parrillo, *The De-Privatization of American Warfare: How the U.S. Government Used, Regulated, and Ultimately Abandoned Privateering in the Nineteenth Century*, 19 YALE J.L. & HUMANITIES 1 (2007)

[86] *Id.*

[87] *See* Evans, *supra* note 2, at 99-100; Cate, *supra* note 14, at 1798 (centralized data collection centers).

Nonetheless, much as in the taxation context, the trend is shifting in favor of government collection. The federal government has become, by far, the biggest collector of health data, holding one billion Medicare claims alone.[88] As I describe above, the FDA, CMS, and ACA programs have begun amassing large quantities of data.[89] State governments also hold substantial amounts of data. Several states require insurance companies to submit to them all the claims data they have collected.[90] Health information is also collected through surveys such as the National Health Interview Survey and the National Health and Nutrition Examination Survey.[91] As I explain in the next Part, for the health information system to operate efficiently and gain legitimacy, the government's role will and should continue to grow.

B. The Normative Case

Society generally, if grudgingly, accepts the necessity of some taxation. Those who treat taxation as morally acceptable offer four primary justifications:[92] (1) paying back to society the benefits one enjoys from living in it; (2) redistributing benefits from the better off to the less well off; (3) participating, in the running of, and expressing allegiance to, one's society as a form of solidarity; and (4) discouraging certain behaviors and encouraging

---

[88] Kristin Madison, *Health Regulators as Data Stewards*, 92 N.C. L. REV. 1605, 1607 (2014).

[89] *See* 21 U.S.C. 355 (k) (3)(B) (2012) (creating the sentinel system to work with "public academic and private entities" to "develop validated methods for the establishment of a postmarket risk identification and analysis system to link and analyze safety data from multiple sources."). *See generally* Madison, *supra* note 88, at 1615.

[90] *Id.* at 1610 n. 24 (discussing allclaims data in state. Laakmann, *supra* note 15, at 42-43 (discussing the FDA).

[91] Madison, *supra* note 88, at 1612.

[92] I deliberately leave the term "justifications" here ambiguous. The justifications at very least form the basis of the obligation of individuals to provide money and information to the State. The same "justifications" may also serve as the basis for government enforcement to ensure that individuals satisfy the obligations. However, one may believe that individuals have no obligation to pay taxes, or one may believe that while individuals have an obligation to pay taxes, it is illegitimate or unjustifiable for the government to *enforce* that obligation. *See* Wertheimer, 54. Whatever the case is, the scope of health information collection will likely be limited by a reader's stance on taxation. A reader who believes there is no obligation to pay taxes will likely see no obligation to provide health information; one who believes that there is such an obligation but no legitimacy or justifiability in enforcing it is likely to believe the same of enforcing health information collection. Finally, of course, one may argue that health information involves a more grievous intrusion, and therefore the two categories are not analogizable. *But see* discussion in Subsection II.A.1 *supra*.

others.[93] The first three of these track popular bioethical justifications for health information collection. Each justification exerts different weight in each collection system, but ultimately all the justifications play a role in both contexts. [94]

1. The Benefit Principle

The benefits principle justifies taxation on the ground that one should pay back to society the benefits one gleans from the public goods that society provides. It is perhaps best reflected in the original reasoning underlying social security legislation. The Act's main architects argued that the form of "insurance" the Act represented was justified by a benefits principle: what one got out of the system was tied to what one put in it.[95] This was not some form of "public relief" funded out of general tax revenues.

There are two different ways to measure "benefits." On the narrower account, benefits include only the direct benefit one personally receives—such as social security payments. The broader account treats all direct or indirect advantages of being a member of society as benefits. It counts the benefits an individual gains from social arrangements, economic conditions, law, and infrastructure, even if she receives no direct payments from society.[96] Laws

---

[93] Scholars mainly emphasize the benefits and redistribution goals, sometimes discussing a third regulatory goal (my fourth goal). *See* Reuven S. Avi-Yonah, *The Three Goals of Taxation*, 60 TAX L. REV. 1, 12 (2006); *see also* Eric Rakowski, *Can Wealth Taxes Be Justified?*, 53 TAX L. REV. 263, 288 (2000); Edward J. McCaffery, *A New Understanding of Tax*, 103 MICH. L. REV. 807, 831 (2005); Daniel N. Shaviro, *Commentary Inequality, Wealth, and Endowment*, 53 TAX L. REV. 397, 400 (2000); James R. Repetti, *Democracy and Opportunity: A New Paradigm in Tax Equity*, 61 VAND. L. REV. 1129, 1135 (2008). Solidarity is a comparatively newer thread in the literature that is cited below.

[94] *See, e.g.*, Hobbs et al., *supra* note 41, at 273, 279-80 (contrasting German and British approaches, each of which, I would argue, emphasize the approaches I lay out above to different degrees).

[95] ROY LUBOVE, THE STRUGGLE FOR SOCIAL SECURITY 174 (1986). Of course, one may always at the end of the day obtain from insurance more or less than what you put in. But the potential benefits relate to one's initial contributions.

[96] The approach also justifies Supreme Court jurisprudence that allows taxing out of state residents for the benefits they get from conducting business within the state. Richard J. Wood, *Supreme Court Jurisprudence of Tax Fairness*, 36 SETON HALL L. REV. 421, 475 (2006). *See also* Compania General de Tabacos de Filipinas v. Collector of Internal Revenue, 275 U.S. 87, 100 (1927) (Holmes, J., dissenting) ("[t]axes are what we pay for civilized society."); Joseph M. Dodge, *Theories of Tax Justice: Ruminations on the Benefit, Partnership, and Ability-to-Pay Principles*, 58 TAX L. REV. 399, 399-400 (2005); RICHARD A. MUSGRAVE, THE THEORY OF PUBLIC FINANCE: A STUDY IN PUBLIC ECONOMY 9-22 (1959). Admittedly while the broader version of the principle claims, not just that a rich individual should pay for the

that allow you to inherit property from your parents, as opposed to escheating all property to the state, fall into this category for example. Those who are poor get fewer benefits from the system overall and therefore owe less to society.

Major aspects of federal information regulatory policy so far implicitly reflect the narrow version of the benefit principle. For example, CMS's program mandates "meaningful use" of Electronic Health Records (EHRs).[97] HHS adopted a "phased approach" to achieve this goal.[98] The first two phases collect information to benefit the patient that provides it. Stage 1 focuses on "electronically capturing health information in a structured format,"[99] to help "disease and medication management" for the patient.[100] Stage 2 focuses on promoting information exchange between the various providers the particular patient relies upon.[101]

The patient personally reaps the benefits of the collection. First, the information the doctor collects is used to treat the patient. Second, providing the information may ultimately redound to the patient's benefit down the line. Later providers will have access to the information regarding previous diagnoses, allergies, drugs, and even DNA profiles the patient may provide to an initial provider. This can help prevent adverse treatment interactions down the road.[102] Further, if the patient becomes unable to communicate or recall previous information, changes from pre-recorded biometric information

---

benefits received, but may need to pay more, examining what is "more" or "less" compared to unmeasurable benefits seem abstract enough that I do not address that question here. *See* Dodge, *supra*, at 412.

[97] Section 1848(o)(2)(A) and § 1886(n)(3)(A) of American Recovery and Reinvestment Act of 2009.

[98] Under final rule, 75 Fed. Reg. 44313 (discussed under definition of meaningful use). *See also* 77 Fed. Reg. 53973 (stage 2 update).

[99] *Id.*

[100] *Id.*

[101] *Id.*

[102] Thus, maintaining active medication and allergy lists are core components of Stage 1, *Meaningful Use Stage 1 Checklist*, HITEC-LA, http://www.hitecla.org/mu_checklist (last accessed Aug. 25, 2015). A recent report from Canada's Infoway program similarly discusses numerous benefits to the specific patient such as reducing duplicate tests. *See* CANADA HEALTH INFOWAY, THE EMERGING BENEFITS OF ELECTRONIC MEDICAL RECORD USE IN COMMUNITY-BASED CARE 130-52 (2013), *available at* https://www.infoway-inforoute.ca/en/component/edocman/resources/reports/benefits-evaluation/1224-the-emerging-benefits-of-electronic-medical-record-use-in-community-based-care-full-report.

such as BMI, weight, blood pressure, or information about past behavioral conditions, such as smoking, can prove important as diagnostic tools.[103]

Importantly, neither stage focuses on passing on information collected from the patient to the rest of society in order for other individuals to benefit.[104] Just like the payer in the social security context, the patient gets a *direct* benefit that is related and proportional to the resources she paid in.

In the scholarly context, many ethicists adopt the broader variety of the benefits principle. They argue that individuals benefit (either now or in the future) from the advanced medical resources society produces when its members contribute to research. Thus, they argue (not without controversy) that these individuals must contribute back to research.[105] Inter alia, this includes providing medical information.

This argument can be extended further. The good health an individual enjoys is not just a result of others' research contributions. As with the case of income, social arrangements, including laws, education, etc., determine one's health.[106] The *value* of that good health, in turn, depends on social arrangements. Some societies penalize bad health more than others.[107] Further, even an individual who owes her health simply to good genes rather

---

[103] *Id.*

[104] The Stage 1 language in passing notes that EHR will allow the "reporting [of] clinical quality measures and public health information," but this is not a primary goal, and is left undeveloped. The vast majority of public reporting checklist items remain optional in both stages. Stage 2 includes an immunization reporting core requirement, but public health surveillance, for example, remains optional.

[105] *See also* Hoffman & Podgursku, *supra* note 3, 125 (2012); Joanna Forsberg et al., *Why Participating in (Certain) Scientific Research is a Moral Duty*, 40 J. MED. ETHICS 325 (2014). People interact in many different common and intersecting zones, including the economy, the government, and the educational system. Everyone benefits from the health of others within each zone, just as everyone benefits, in theory, from universal, compulsory education. The presumption is that society is better served, in ways that might be impossible to delineate . C. D. Herrera, *Universal Compulsory Service in Medical Research*, 24 THEORETICAL MED. 215, 220 (2003). In addition, once we view health as a social product, it becomes less important whether we feel that we benefit, as individuals, from medical research. The broader view has us moving away from the individualistic thinking that encourages us to care only whether we or those we know can remain healthy. A viable system of universal service would be based on the idea that there is a social good in living amongst those who are in fairly good health. *Id.* at 219; *see also* Donald Willison, *Privacy and the Secondary use of Data for Health Research*, 8 J. HEALTH SERV. RES. & POL'Y S1, S17, S19 (2003) (research should benefit the community from whom it is collected).

[106] *See* DANIELS, *supra* note 44, at 34 (explaining shortcomings in previous work).

[107] *Id.* at 43 (defending health as a basic Rawlsian social good because it offers opportunity); *See* Chai R. Feldblum, *Rectifying the Tilt: Equality Lessons from Religion, Disability, Sexual Orientation, and Transgender*, 54 ME. L. REV. 159, 184-85 (2002).

than social arrangements benefits from society's protection of the health of others. Law professors benefit from the good health of colleagues who can read and comment on their work, and law review editors who source cite and publish articles. And of course, ill health of those around you can place a burden on you: "[o]ne person's malady can harm families, workplaces, clubs, churches, and sometimes entire communities."[108] Social programs like Medicaid are invaluable for maintaining this productivity and evading burdens.[109] Thus, even healthy individuals who do not directly consume health benefits owe society under the broader benefit principle—and can contribute by providing, inter alia, their health information.

2. The Redistribution Principle

Redistribution involves asking those who are better off to pay more into society.[110] The better-off may benefit to some extent, when their taxes go towards schools, roads, and national defense. However, the scheme is designed to help those who are worse-off.

Scholars offer various justifications for redistribution.[111] Under a utilitarian approach, the poor value a marginal improvement in the outcomes they experience more than the rich value a marginal loss. Hence, redistribution helps improve overall societal utility. Another consequentialist goal is to ensure the proper functioning of democracy. If some individuals experience outcomes below a certain threshold, they will be unable to participate in political life. In a non-consequentialist vein, redistribution is part of a general duty of beneficence to help those in dire straits—analogous to a duty to rescue.[112] Independent of that duty, inequality may raise

---

[108] William M. Sage, *Solidarity: Unfashionable but still American*, in CONNECTING AMERICAN VALUES WITH HEALTH REFORM (Hastings Center ed., 2009).

[109] *See, e.g.*, Jenkins & Konecny, Nebraska Medicaid Expansion 27-38 (2015) (explaining and reviewing literature showing that Medicaid expansion boosts incomes and productivity) *available at* http://www.nebraskahospitals.org/file_download/inline/9eb5a4d7-8725-4385-959f-0f404c895128.

[110] As I note above, a version of the benefits principle by itself supports some version of redistribution—because those in better circumstances benefit the most from the system, they must pay more to the system. Thus, many think of the redistributive and benefits goal as overlapping. Avi-Yonah, *supra* note 93, at 12; C. Eugene Steuerle, *And Equal (Tax) Justice for All?, in* TAX JUSTICE: THE ONGOING DEBATE 233 (Joseph Thorndike & Dennis Ventry Jr. eds., 2002); Yoseph Edrey, *Constitutional Review and Tax Law: An Analytical Framework*, 56 AM. U. L. REV. 1187, 1228 (2007).

[111] *See, e.g.,* Dodge, *supra* note 96, at 430.

[112] Harris, *supra* note 57.

particular concerns. Well-known economist and tax theorist, Henry Simons objected to inequality on the basis of an "ethical or aesthetic judgment [to] the prevailing distribution of wealth and income [which] reveals a degree (and/or kind) of inequality which is distinctly evil or unlovely."[113]

In the health context, bioethicists who support a duty to participate in research have based this duty in part on a duty to aid those who are worse-off. They impose this duty notwithstanding the burdens it imposes on the (often times, better-off) research participants.[114]

In the informational research context, research involves risks to its subjects.[115] Nonetheless, the bioethical argument requires individuals better off in the health context—that is, healthy individuals—to provide information in spite of these risks.[116] The information will then be used to improve medical knowledge, to improve the outcomes that unwell individuals experience.

3. The Solidarity Principle

According to a new vein of legal scholarship,[117] paying tax is a way of performing and exhibiting solidarity with a particular group. Paying taxes shows civil and communal responsibility, even patriotism. Payment allows individuals to symbolically join together in carrying out a common activity, thus showing they are part of a single community. It also shows trust and commitment to the common mission, by delegating spending decisions to the government.[118] Accordingly, paying taxes ultimately denotes identity as a member of a particular community.[119]

---

[113] HENRY SIMONS, PERSONAL INCOME TAXATION 18-19 (1938). *See also* WALTER BLUM & HARRY KALVEN, THE UNEASY CASE FOR PROGRESSIVE TAXATION (1952).

[114] *See supra* note 57.

[115] Brassington, *supra* note 57.

[116] David Orentlicher, *Making Research a Requirement of Treatment: Why we Should Sometimes let Doctors Pressure Patients to Participate in Research*, 35 HASTINGS CTR. REP. 20 (2005) suggests that only sick people experience burdens. *But cf.* Franklin Miller, *Ethical Issues in Research with Healthy Volunteers*, 6 CLINICAL PHARMACOLOGY & THERAPEUTICS 513 (2003); Jonathan Perlin & Joel Kupersmith, *Information Technology and the Inferential Gap*, 26 HEALTH AFF. w192 (2007) (noting that RCTs often will not include individuals who are extremely ill).

[117] Ajay K. Mehrotra, *The Price of Conflict: War, Taxes, and the Politics of Fiscal Citizenship*, 108 MICH. L. REV 1053 (2010).

[118] Lawrence Zelenak, *Taxing Endowment*, 55 DUKE L.J. 1145, 1153 (2006).

[119] Erik Christensen, *Biobanks and our Common Good*, *in* THE ETHICS OF RESEARCH BIOBANKING, *supra* note 73, at 101, makes this argument, but confuses it with various other

This taxation justification is particularly exploited during wartime. Through regressive consumption taxes and graduated income taxes, Lincoln sought to send the message that "all Northern citizens were supporting the war equally," and "employ[ed] the rhetoric of patriotism and shared sacrifice." Similarly, as historian Carol Jones argues, the dramatic expansion of the tax base during World War II could not rely on redistributivist arguments given that it negatively affected the less well-off. Rather, the emphasis was on that of a joint effort by all citizens, to share the cost of war.[120]

The bioethics literature has increasingly argued that solidarity imposes a duty to aid in research. The goals of health information research are inherently communal. Medical knowledge is "a social product…, i.e., [] its production involves the cooperation of many individuals, and…these individuals enter into specific forms of social relations in producing it."[121] Its goals cannot be achieved without our coming together.[122] Indeed, the need for communal involvement in the health information context is far greater in the taxation context. Tax renegades will lower the government's revenue stream, but the government will be able to carry out some work. Information renegades could destroy the statistical robustness of entire studies.[123]

---

redistributive and benefits based principles. *See* Jennings, *Public Health and Civic Republicanism, supra* note 55, at 20 for a recent account of solidarity ("mutuality of care and a public expression of reconition and concern"). Jennings argues that solidarity is characterized by three "relational dimension[s]", the first, "standing up *for*" the community, the next, "standing up *with*" the community, and the third, "standing up *as*" a member of the community. The benefits principle is a distinct matter—even if you live abroad and get few of the benefits from living within the country, the solidarity principle requires you to pay taxes, as many American expatriates must, in fact, do. Michael S. Kirsch, *Taxing Citizens In A Global Economy*, 82 N.Y.U. L. REV. 443, 445 (2007) ("With certain exceptions, the United States taxes the worldwide income of its citizens regardless of whether the citizen lives in the United States or abroad.").

[120] Carol Jones, *in* FUNDING THE MODERN AMERICAN STATE, 1941-1995 at 42, 114 (W. Elliot Brownlee ed., 1996); *see also* Mehrotra, *supra* note 117. Indeed, *contra* Avi-Yonah, *supra* note 93, at 12 who analogizes paying taxes to language, history and culture, these other artifacts are far more fragmenting in American society than tax payment, and are poor analogies.

[121] M.W. Wartofsky, *Medical Knowledge as a Social Product: Rights, Risks, and Responsibilities, in* NEW KNOWLEDGE IN THE BIOMEDICAL SCIENCES: SOME MORAL IMPLICATIONS OF ITS ACQUISITION POSSESSION AND USE 113 (W.B. Bondeson et al. eds, 1982).

[122] Bartha Maria Knoppers & Ruth Chadwick, *Human Genetic Research: Emerging Trends in Ethics*, 6 NATURE REVIEWS GENETICS 75 (2005).

[123] *Supra* note 65.

Solidarity values have motivated information collection in real world contexts. Providing health information can shape identity.[124] Studies have shown that providing information regarding a certain condition reinforces individuals' commitment to and identification with the relevant patient community. Quite apart from potential benefits that may accrue from future research, individuals share information as a way to build ties and solidarity to form "regimes of biosociality."[125]

This group-expressive act of health information collection can occur at a national level. Other countries such as Iceland and Britain have emphasized how their biobanks represent national uniqueness or diversity, a source of national pride, and how their research will further important national goals.[126] Thus, in creating Britain's national biobank, participants suggested that the fruits of their research made them "smile."[127] They felt they were "part" of a collective endeavor "that was very worthwhile."[128]

Solidarity values may have been part of the inspiration for domestic policies as well. PCORNet and PMI seek to make their data broadly available for all researchers. Similarly, Stage 3 of CMS's meaningful use program is best justified on the grounds of solidarity. Stage 3 (per proposed rule at the time of writing) incorporates Stage 1 and Stage 2 goals. But Stage 3 also promotes collecting mass data into centralized locations such as public health agencies and general clinical registries. [129] Although this can be justified on

---

[124] *See* Subsection I.A.2.

[125] Sandra See-Jin Lee, *Social Networking in the Age of Personal Genomics*, 3 ST. LOUIS U. J. HEALTH L. & POL'Y 41 (2009). Though under the existing system, those who can provide this information tend to be upper middle class or richer, educated, and white.

[126] Kadri Simm, *The Concepts of Common Good and Public Interest: From Plato to Biobanking*, 20 CAMBRIDGE Q. HEALTHCARE ETHICS 554, 558-60 (2011); Gisli Palsson, *The Rise and Fall of a Biobank: the Case of Iceland*, *in* BIOBANKS: GOVERNANCE IN COMPARATIVE PERSPECTIVE 41, 45-46 (Herbert Gottweis & Alan Petersen eds., 2008). In yet another example, Estonia found that by shifting the justification to emphasize national solidarity helped increase support for the biobank. Simm, *supra*, at 560

[127] Hobbs, *supra* note 41, 278. Busby & Martin, *supra* note 41.

[128] *See also* E. Haimes & M. Whong-Barr*, Levels and Styles of Participation in Genetic Databases: A Case Study of the North Cumbria Community Genetics Project., in* GENETIC DATABASES: SOCIO-ETHICAL ISSUES IN THE COLLECTION OF DNA (Richard Tutton & Oonagh Corrigan eds., 2004) (noting some individuals' allegiance to the region as a motivation for giving blood).

[129] The goal is to "increase[] focus on the importance of the ongoing lines of communication that should exist between providers and public health agencies (PHAs) or…providers and clinical data registries (CDRs)." Medicare and Medicaid Programs; Electronic Health Record Incentive Program-Stage 3, 80 Fed. Reg. 16,732 (March 30, 2015). Unlike Stage 2, where non-immunization reporting was optional (and limited), including only cancer registries and

multiple grounds,[130] I read the primary justification of Stage 3 as one of solidarity. The various objectives of Stage 3 are united under that of "improving population health."[131] Stage 3 therefore conceptualizes a unified community that it seeks to benefit. It creates mechanisms that bring individuals together, in solidarity with each other, such as dedicated central repositories and registries.

It bears noting, however, that for many bioethicists, the solidarity values at stake implicate an "imagined community" beyond the nation, to encompass all of humanity. While individual genes and other biological conditions create distinguishable profiles that represent the individual, the basic structure of the genome represents a picture of humanity itself, representing us at "the level of a species."[132] The ideal type human, its internal structure, and its failings are all represented by the composite that the medical research we bring together creates. In this way, unlike the case of financial pooling, the pooling of health information is not a *means* to achieve public goods, but is the public—or a representation of the public—itself.[133] By contributing towards this effort, the individual herself—not just her interests—begins to meld with the community.

4. The Regulatory Principle

---

specialized registries. Medicare and Medicaid Programs; Electronic Health Record Incentive Program-Stage 2, 77 Fed. Reg. 53,968 (Sept. 4, 2012) However, consistent with a general reluctance to preempt local privacy law, Stage 3 continues to provide exemptions in cases where localities prohibit certain forms of reporting.

[130] For example, some may argue, Stage 3 is simply a broader application of the benefits principle. Stages 1 and 2 represented a narrow version of the benefits principle—patients only provide information to benefit themselves down the line. Those Stages did not, as a general matter, emphasize reporting individual patient information to benefit the collective. Stage 3 represents a broader version of the benefits principle. Individuals provide information to the collective as a form of payback for the indirect benefits they derive from the medical knowledge that the collective generates. And I do not deny this as a possibility—indeed, as I note in the introduction to this Part, the health information collection system like the taxation system is and should be quite comfortable with multiple and competing justifications.

[131] 77 Fed. Reg. 53973.

[132] Knoppers & Chadwick, *supra* note 122, at 77.

[133] *Id.* at 208 ("Genetics has demonstrated our profound connectedness to our siblings, parents, other humans and even animals, raising questions about the very possibility of drawing a line between what is mine and what is the others."). Indeed, one could go further and argue that such genetic representation makes it hard to recognize the line between what is uniquely me, and the rest of humanity; our very understanding of autonomy may be in question.

The final reason for revenue taxation is to force individuals to internalize the costs of unhealthy behavior to deter that behavior. For example, we may tax cigarettes or fine a failure to hold insurance.[134] We may also collect information to deter undesirable behavior. In other work, I argue that surveillance and information collection can stigmatize and impose costs on individuals.[135] Certain information is collected about the individual, which in turn imposes substantial social cost on her. Because this justification of taxation is not directly germane to the collective pooling of information for research, and because it has not figured prominently in the literature, I do not examine it in detail.

III. CONCEPTUALIZING AND ORDERING THE SYSTEM

Drawing an analogy between information and tax collection helps reconceptualize and provide guidance for the information system. But is analogical reasoning the best approach for this? An alternative deductive approach would entail identifying some appropriate normative system—Rawlsianism, utilitarianism, and the like, from which we would derive lessons designed for the health information context.[136] Prominent bioethicists have done just this.[137]

I believe that analogical reasoning offers at least five benefits. First, my goal here is not just to derive ethical principles but to offer another paradigm as a lens for the new effort. The taxation analogy reshapes the way we see the entire collection enterprise—as one system, rather than multiple programs.

---

[134] *See, e.g.*, David A. Strauss, *Commerce Clause Revisionism and the Affordable Care Act*, 2012 SUP. CT. REV. 1, 8 (2012) ("The mandate 'creates a financial incentive (by means of a tax penalty) for uninsured participants in the health care market to internalize their own risks and costs, rather than externalizing them to others.'").

Some version of the benefits principle may justify internalization of costs: individuals must pay for the costs they will later impose on society. But one may impose a tax that theoretically exceeds social costs to further discourage the behavior. The regulatory principle is the primary justification for a tax when the main aim of the tax is deterrence—in such cases, the fact that society obtains revenue as repayment for the costly behavior is incidental.

[135] Konnoth, *supra* note 46; Avi-Yonah, *supra* note 93, at 25.

[136] *See* Cass Sunstein, *On Analogical Reasoning*, 106 HARV. L REV. 741, 753 (1993).

[137] Daniels, *supra* note __; Jennings, *supra* note __.

Second, individuals are familiar with income taxation. An analogy based on it will make the currently obscure health information system, seem more approachable and comprehensible.[138]

Third, the tax analogy lays out the scope of the problem. The reach of the new health information collection system is broad. Even if we can derive reasonable answers, how can we know in advance what questions to ask? The tax model has already engaged with similar issues over a long period of time and will give us some sense of the detail and scope with which we must address problems.

Fourth, analogical reasoning is ideal where there are multiple, sometimes conflicting justifications, for a particular practice, as here. Deducing moral principles in such cases is difficult—individuals may not agree on the underlying justifications of either taxation or health information collection.[139] What combination of the benefits, redistribution, or solidarity principles undergird each system? Do all justifications have a role to play? With analogical reasoning, we can just point to similarities between the systems. Based on this, we can argue that the health information collection system should consider following some or all of the prescriptions of the taxation system, without claiming the superiority of any one of these normative visions.[140]

Finally, and perhaps most importantly, at the high level of generality at which they are discussed here, the principles of justice that undergird taxation have already been largely publicly vetted and explored in a well-established literature. Although individuals are not favorably disposed towards taxation, there is (frequently unenthusiastic) overlapping consensus as to its principles from multiple constituencies with different approaches to questions of justice and morality. Modern mainstream controversies involve

---

[138] GEORGE LAKOFF & MARK JOHNSON, METAPHORS WE LIVE BY 6 (1979).

[139] Sunstein, *supra* note 136, at 755. Analogical reasoning works best in cases involving "undertheorized practices." In these cases individuals are unable to agree on a general theory to explain the practice, or "too many factors are relevant, and too many variations are possible, to allow a general formulation adequately to capture the range of right results in the cases.")

[140] This is particularly germane in the health research ethics context where individuals are often unable to agree on justifications for reaserch. As Sunstein puts it, analogy "allow[s] people unable to reach anything like an accord on general principles to agree on particular outcomes…An overlapping consensus is often possible on the view that case A is relevantly similar to case B even if those who join the consensus could not decide as between utilitarianism or Kantianism, or come to agreement on the appropriate role of religion in society." *Id.* at 745.

alterations at the margins in ways that do not affect my reliance on its basic values. I need not assume agreement on sometimes controversial foundational theories of justice to make my case.

Accordingly, in what follows, I argue that the analogy helps reconceptualize the health information collection system as a singular system. This lays the groundwork for important substantive and procedural justice values that we should at least consider importing into the health information context. In the long run, this will begin a discussion about the principles of justice in the health information context in their own right, without any dependence on tax frames.

A. Systematic Perspective

Perhaps the biggest problem that faces the modern health information effort is that policymakers and scholars understand it through the lens of traditional health information collection. 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….In essence, the laws have developed by putting out fires, without comprehensive planning for modern public health problems."[141]

Because of this, modern collection also treats each act of collection as a standalone enterprise. [142] Considerations advanced with respect to one health collection project, such as increasing providers that are part of the CMS's meaningful use program, may prove completely inapplicable with respect to other health information projects such as the FDA's post-market

---

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

[142] 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. But there is no indication that policy is moving in that direction, nor any guarantee that this will increase information access. *Compare, e.g.*, Mark A. Hall, *Property, Privacy, and the Pursuit of Interconnected Electronic Medical Records*, 95 IOWA L. REV. 631 (2010); *with* ALAN F. WESTIN, PRIVACY AND FREEDOM 324-25 (1967); *see also* Edward J. Janger, *Privacy Property, Information Costs, and the Anticommons*, 64 HASTINGS L.J. 899 (2003) (suggesting that information flow will reduce by giving individuals more control). Further, it would be nigh impossible to stabilize data sources into a reliable health information source. Further, problems regarding bias, *supra* note ___, will remain. Minorities are less willing to provide information, and other individuals minds may change as they grow older, leading to unstable information sources, or too much information with respect to some age groups and not enough with respect to others.

drug surveillance. This has prevented us from developing any systemwide principles or goals with which to evaluate health collection efforts at a more general level.

The account in the previous Part identifies common characteristics of various health information collection programs. But identifying common characteristics among multiple programs is not the same as seeing them as part of a single system and guided by a common set of core interdependent principles. Indeed, the fact that different programs may rely on multiple justifications—sometimes benefits-related, sometimes redistributive, and sometimes solidarity-based—may further undermine our ability to conceptualize health information collection as a single system.

Taxation, however, also relies on multiple justifications, but is, nonetheless, considered part of a single system. Tax historian Elliott Brownlee explains that tax policy and theory is inherently "pluralis[tic]."[143] Our understanding is that all kinds of income taxation involves collecting resources from individuals (citizens, residents, and some others); that it has limits; that it serves the public good; and that it sometimes operates as insurance, as redistribution, as regulatory incentive, and as a show of solidarity. Indeed, specific policies are able to attract support from varied constituencies because of the pluralistic reasoning they may invoke. In spite of this, taxation is understood as part of a single, wide ranging system in service of a certain set of goals.

To be sure, given taxation's conceptual pluralism, one cannot predict policy outcomes in a given circumstance. But that does not undermine our understanding of taxation as a *system*. Within a particular location of the system, certain policy judgments may come out differently depending on how the discourses I outline are balanced. But each of these discourses permeates the tax system and makes it cohesive, if not coherent. At every point within this system, these arguments are logically, if not politically or culturally, relevant to a given policy choice. These multiple theoretical underpinnings ultimately constrain the arguments that can legitimately be advanced in any one particular revenue policy context.

The taxation analogy opens up the possibility for modern health information collection to be understood as part of a single system. Like the taxation system, multiple normative accounts play a role in the system and may be deployed differently across different contexts. As in the taxation

---

[143] W. Elliott Brownlee, *Reflections on the History of Taxation*, *in* FUNDING THE MODERN AMERICAN STATE, *supra* note 131, at 3, 13.

context, this "conceptual pluralism" allows specific policies to appeal to different constituencies and helps refine policy decisionmaking.[144] Individuals will invoke each discourse depending on the specific policy agenda they seek to advance, but the overall goals of the system, its tools, and policy levers are all part of a single effort.

The conceptual assistance that the tax analogy therefore offers is partly an aesthetic one. It helps unite the disparate jigsawed elements of health information collection, present, and future. This does not eliminate questions, but allows us to approach them from the vantage point of a potentially unified frame. We may keep in mind "the big picture," as one set of bioethicists puts it, even as we examine detailed questions.[145]

B. Normative Principles

Understanding the health information enterprise as comprising a single system is a useful precursor to applying principles of substantive and procedural justice to govern its operation. The tax analogy proves useful in developing these principles as well.

In his foundational work on taxation theory, Adam Smith argued that taxation should conform to four essential principles: tax equity (as understood today, both horizontal and vertical), clarity and certainty of payment, ease of collection, and production of value that equates to the amount of the collected tax.[146] These principles reflect three underlying concerns: (1) substantive justice that ensures that the benefits and burdens of collection are fairly distributed; (2) procedural justice that requires that the means and methods of collection are transparent and legitimate; and (3) efficiency that ensures that the collection is carried out easily and with minimal waste.

Applying similar values to health information collection would be unworkable if they massively contradicted existing, entrenched policy or moral approaches. However, the analogy provides answers that systematize

---

[144] Conceptual pluralism plays a role in shoring up support for other healthcare systems. *See* Allison Hoffman, *Three Models of Health Insurance*, 159 U. PENN L. REV. 1873, 1882 (2011) (explaining how the conceptual pluralism that characterizes the justifications for the ACA shores up its support and helps refine ACA programs); *see also* Sunstein, *supra* note 136, at 747 (explaining that conceptual pluralism is necessary for wide ranging programs to have relevance to multiple areas of our lives and to engage a large variety of people).

[145] Margit Sutrop & Kadri Simm, *Public and Private Interests in the Genomic Era, in* 4 ETHICS LAW AND SOCIETY 205, 213 (2009).

[146] ADAM SMITH, THE WEALTH OF NATIONS 888-89 (Edwin Cannan ed., Random House 2d ed. 2000) (1776).

rather than revolutionize. In most areas, it fills a void, providing guidance where there was no direction before. In other areas, where there is existing theory and practice, the solution the analogy provides turns out to be one towards which scholarship or practices regarding health collection have already been moving. The taxation analogy in these latter cases nudges these advancements slightly forward and frame them as part of a larger system, without demanding a sea-change.

Ultimately, there are three takeaways from this Section. First, we must calibrate the burdens and benefits of privacy risk distribution based on an individual's level of welfare. Second, we must move away from a contractual/private law, individual informed consent model of collection to a public law, democratic consent model of collection. And third, the government should take the lead in collection for efficiency purposes.

The values I lay out here must be translated into actual processes for fair and legitimate administration of information collection, agency bodies to oversee the processes, and products and services that allow access to benefits. But my task here is limited to borrowing values from the taxation context and showing how they may be modified to apply to the health information context. Apart from a few demonstrative examples, a systematic exposition of the implementation of these values must await future work.

1. Substantive Justice: Fairness

A virtue of the analogy is that it brings questions of substantive justice into sharper focus. To be sure, like the health information apparatus, taxation seeks to optimize and streamline revenue collection. through administrative organization, partnerships, and technological improvement. But taxation is also occupied with questions of fairness. These fairness considerations involve many principles, some of which are in tension with and limit each other. The dominant fairness concern is that of equity.

Taxation's architects put serious thought into how the *burdens* of taxation should be distributed in an equitable manner given individuals' welfare-levels. The threshold question is what set of criteria determines the welfare of an individual. Income is a key metric of welfare, but not the only one.[147] Tax scholar Eugene Steurle notes that the tax system also calibrates, or should calibrate, its measure of an individual's welfare using her health (measured through her health expenses or disability), geographical location

[147] Steuerle, *supra* note 110, at 274-76].

which may affect the purchasing power of the individual's income, family size,[148] age, etc. We treat individuals who have the same incomes, but meaningfully differ with respect to the other characteristics as differentially situated. Debates remain as to what appropriate characteristics are, and the weight they should be given in measuring welfare.[149]

The manner in which we distribute the *benefits* of taxation also reflects welfare metrics. Although a person's welfare level is not the *sole* determinant, the state often distributes health, food, and income depending on an individual's finances, age, health, or family size, and frequently uses detailed measures. For example, whether an individual qualifies for health-related benefits is determined by a detailed set of health measures subject to several levels of administrative and judicial review.[150]

Tax justice prescribes two principles that determine how to distribute burdens and benefits based on welfare. Under horizontal equity, those who

---

[148] Peter Lambert & Shlomo Yitzhaki, *Income Tax Credits and Exemptions*, 13 EUR. J. POL. ECON. 343, 344 (1997), using income and family size. Even if one just used income, what income metric should be used? Alan Auerbach & Kevin Hassett, A New Measure of Horizontal Equity 2, 4 (NBER Working Paper No. 7035, 1999), *available at* http://www.nber.org/papers/w7035.pdf; *see generally* Steurele, *supra* note 110, at 273-78

[149] *See, e.g.,* Daniel Markovits, *Luck Egalitarianism and Political Solidarity*, 9 THEORETICAL INQUIRIES LAW 271 (2007). For example, even with as fundamental a measure as income, some argue that *potential* rather than *actual* income is the better measure. Some argue 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, they 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 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. Steurle, *supra* note 110, at 258. Accordingly, we rely on actual rather than potential income to assess taxes.

Interestingly, as in 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 his 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 unethical (though perhaps not impractical) to increase burdens on those with ill-health. Finally, measuring actual health is feasible; measuring potential health under current technology, is not.

[150] *See* Social Security Administration, Disability Evaluation Under Social Security, https://www.ssa.gov/disability/professionals/bluebook/ (last visited, Apr. 26, 2016) (providing a listing of the numerous conditions social security covers).

are similarly well-off should be similarly treated. Vertical equity may require adjusting burdens and benefits across those with different levels of welfare.[151] Although there remains dispute as to how to make this judgement, those with higher levels of welfare end up net losers in the current tax system. [152] They usually suffer the highest financial burdens imposed through income collection, and receive the fewest benefits either in income or in kind. As discussed in the previous Part, we justify this approach under the broader benefit principle—those who have experienced greater welfare owe society more—and the redistributive principle—the mandate to help individuals worse off than oneself.

Since the benefit and redistributive principles also apply in the health information context, the welfare based model of distribution should apply. The measure of welfare will remain similar or the same: income is a primary determinant, along with health status family size, or age. The key difference is that the burdens consist, not of lost *income*, but rather, *privacy risk* due to collected information. The benefits of health information collection involve a narrower range of resources involving health research and care delivery.

Vertical equity of burdens requires exempting the worse off from privacy risk to the degree possible, and shifting risk to the better off. As in the tax context, such adjustment is accounted for by the broad benefit principle, under which the better off should pay more back to society, and the redistributive mandate to help out the worse off.

In practice, this brings into question key aspects of the health information collection apparatus which benefits the better off and harms the worse off. For example, the system subjects the poor to the greatest privacy risk, and exempts the rich who can pay out of pocket. The CMS meaningful use program promotes the collection of Medicare and Medicaid data which concerns recipients who are largely poor or elderly. Deidentified Medicare and Medicaid data has historically been made publicly available for research.[153] Most individuals not on Medicare or Medicaid still experience privacy risk. Numerous payer and provider networks require collection and analysis of data. Many individuals cannot choose to avoid these providers or payers, since they cannot pick which health insurance they will receive. Richer individuals, by contrast, can pick any doctor they want, and pay out of pocket, thus escaping information collection requirements.

---

[151] Steuerle, *supra* note 118, at 257.

[152] For an overview of this long standing controversy, *see, e.g.,* Barabara Fried, *Why Proportionate Taxation, in* TAX JUSTICE, *supra* note 110, at 149.

[153] *See supra* note 34.

Fair risk distribution requires programs that expand beyond Medicare and Medicaid based information collection. The government should explore ways to ensure that all providers report information on all patients, even richer ones choosing to pay out of pocket. This can be achieved in various ways. For example, we could create a system that can access data distributed across private entities. The Sentinel system offers a nascent, primitive model for this. Or the federal government could require states to make provider reporting a condition of licensing. Each of these approaches are subject to numerous legal and practical considerations which are beyond the scope of this Article. Suffice it to say, however, that federal information collection mandates have not yet been subject to any serious legal challenge.[154]

Health is another measure of welfare. Yet, the existing system disproportionately allows the healthy to escape collection requirements more than the sick because they engage with healthcare differently than sick individuals. To receive care, sick individuals must go to medical providers who are part of the information apparatus. But healthy individuals enjoy "loopholes": they may use personal fitness devices,[155] or rely on checkups that are part of employee wellness programs, which are not always subject to similar collection requirements.

However, the information of health individuals can be extremely useful. Such data provides controls for research involving illness. It can give researchers a sense of the range of normal variation, to help determine at what point certain conditions become pathological. It can help examine innocuous variations of a particular gene or virus,[156] or asymptomatic versions of certain conditions.[157]

---

[154] Large scale information collection mandates by the federal government have not been subjected to serious challenge. However, one could imagine reporting mandates being subject to First Amendment or due process takings challenges. *Cf.* Sorrell v. IMS Health, 131 S.Ct. 2653 (2011).

[155] Nathan Cortez, *The Mobile Health Revolution*, 47 U.C. DAVIS L. REV. 1173 (2014).

[156] *See* Adeline R. Whitney et .al, *Individuality and Variation in Gene Expression Patterns in Human Blood*, 100 PROCEEDINGS NAT'L ACAD. OF SCI. 1896 (2003).

[157] Admittedly data of sicker patients is, as of now, still understood as more useful. Thus, records of sick patients contain more information than healthy patients. Nicole G. Weiskopf, *Sick Patients Have More Data: The Non-Random Completeness of Electronic Health Records*, 2013 AMIA SUMMITS ON TRANSLATIONAL SCI. PROC. 1472, 1476 (2013) ("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 trying to understand what sickness itself is. Hans K. Meier-Ewert et al. *Absence of Diurnal Variation of C-Reactive Protein Concentrations in Healthy Human Subjects*, 47 CLINICAL CHEMISTRY 426 (2001). *See generally* UTAH DEP'T OF

Accordingly, fair distribution requires engaging healthy individuals within the information structure by making a concerted effort to collect their information. Mandates for information collection can extend to contexts in which healthy individuals use health products—fitness devices or wellness programs. This reporting should in no way reduce privacy protections from unscrupulous actors such as discriminatory employers. [158]

Beyond changes in how we collect data, we may even change how we analyze the data to shift privacy risk based on welfare-levels. Depending on the study and the relevant statistical method, we need include only a certain number of patients to yield the appropriate statistical confidence.[159] Querying and transmitting data for a study may increase the risk of reidentification or breach. We should therefore use the data of individuals with less hardship. Thus, if we are reasonably sure that the health data regarding a particular condition is unbiased across income levels, we should use the data primarily of those with higher income. Individuals whose data has been involved in multiple studies may suffer a higher risk of reidentification than others, so their data should be exempted from future studies at some point if possible.[160]

Finally, we should ensure that the distribution of benefits follows welfare-based considerations as well. Research and information collection priorities should focus on aiding the worse-off and marginalized

---

HEALTH, MAKING CENTS OF UTAH'S HEALTHY POPULATION. UTAH ATLAS OF HEALTH CARE (2010) (explaining Utah's effort to collect information about healthy individuals).

[158] *Id.* 1183 ("[O]ne app…pull[s] data from hundreds of different types of patient monitors."); Konnoth, *supra* note 66.

[159] *See* DEPARTMENT OF HEALTH AND HUMAN SERVICES, REPORT TO CONGRESS 21-22 (Oct. 2014), available at http://www.healthit.gov/sites/default/files/rtc_adoption_and_exchange9302014.pdf; PMI Report, *supra* note 16, at 22 (explaining statistical power requirements in nested unmatched case-control design, that limits the number of people required to achieve desired levels of confidence in caste control studies).

[160] *See* Comment Letter from Electronic Frontier Foundation to The Department of Health and Human Services, 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, Jan. 6, 2016, at 5 ("[I]ndividual tests can produce measurements that—when aggregated—can lead to identification.").

communities.[161] Many government programs already adopt this goal, but there is some ways to go.[162]

The vision of equity I describe above will probably not go unchallenged. As in the tax context, many may advocate a narrower version of the benefit principle, that is, that individuals should benefit depending on how much they put in: social security in return for taxes; clinical developments in return for information. Thus, they may argue, it is unfair to collect information from healthy individuals, simply to use them as controls. And, to some extent, they will succeed in limiting the shift of privacy risk from the worse to the better off.

Yet, the current system does not abide by even the narrower version of the benefit principle. A richer individual in today's system will probably directly benefit from the research carried out using data collected from poorer individuals. Thus, even those who support a narrower version of the benefit principle should support some degree of risk redistribution.

Finally, any version of the benefit principle suggests that all individuals who contribute to research should be equally eligible to benefit from the fruits of the resulting public goods. Accordingly, we may conclude that all individuals who participate in research should have equal access to the treatments that the research produces. We therefore should ensure access to payment mechanisms,[163] and invest in technologies to reduce barriers to health.[164]

Equity, is, of course, not the only principle that determines what fairness is. Other values matter, and may limit the reach of equity. For example, we take it for granted that there should be some moral, constitutional, or efficient limit on the collection of revenue. We sometimes exempt certain sources of income from taxation, even if that does not promote

---

[161] Podgursky & Hoffman, *Improving Health Care Outcomes*, *supra* note 5, at 212.

[162] *See* George R. Brown & Kenneth T. Jones, *Incidence of Breast Cancer in a Cohort of 5,135 Transgender Veterans*, 149 BREAST CANCER RES. & TREATMENT 191, 192 (2015); INTEGRATING, *supra* note 4, at 27 ("[W]e know that low-frequency events are important to study, but these demand larger pools of data. We also have strong needs for geographic and demographic diversity.").

[163] To be sure, even in the revenue collection system we exclude some individuals who provide revenue, such as undocumented immigrants, from the benefits of the system to which they contribute. Some may argue that similar reasons (whatever they are) may counsel excluding certain groups from access to health goods to which they contribute. However, as in the revenue system, the presumption should be that individuals have access to the public goods produced partially through their contributions.

[164] Elizabeth Pendo, *Reducing Disparities Through Health Care Reform: Disability and Accessible Medical Equipment*, 2010 UTAH L. REV. 1057 (2010).

equity. These limits are often developed through negotiation in democratic politics. Similarly, we may decide not to collect certain information. For example, some or all information from non-clinical sources such as wearables or grocery cards may be treated as presumptively non-reportable. Other considerations include the nature of the condition, or the existence of social stigma.[165] STDs, for example, are often non-reportable to health exchanges.[166] These decisions should also be developed in the crucible of electoral politics as I develop further in the next Subsection.

Ultimately, choosing between values or projects will involve tradeoffs. Policymakers will have to choose how much redistribution is appropriate, whether resources should be spent on projects that will make a slight improvement for a large number of people, or those that will produce a large improvement for a small number of people. Bioethicists have considered how research priorities should be set—some, for example, have suggested rationing healthcare for the elderly, and focusing more on children and working adults.[167] Congress, however, has forbidden such calculations in certain contexts.[168] These are all difficult, context-based calculations and tradeoffs that must be made at various points across the system, through the process I describe in the next Subsection.[169]

---

[165] Arguably, traditional health information collection practices already represents a precursor to modern collection, except that it is greatly limited in scope. As we move forward, we can take lessons from other countries, which are far ahead of the United States in certain aspects of health information management. Iceland mandates the collection of information regarding health, medical treatment, lifestyles, social circumstances, employment and family.[165] Similarly, Finland collects information, which it stores for upto 100 years (images upto 20 years), in a central archive. Pekka Ruotsalainen et al., *Sharing and Management of EHR Data through a National Archive: Experiences from Finland*, 43 WORLD HOSP. & HEALTH SERV. 38 (2007); *see also* Rodwin, *supra* note 210, at 615-16 arguing that all drug dispensers should report detailed data regarding the dispensing. Further, providers should submit the same patient information to DHHS as they do to third-party payers when seeking payment, deidentified patient profile data, etc.

[166] *See* Konnoth, *supra* note 46 (discussing HIV exceptionalism); Thomas Murray, *Genetic, Exceptionalism and "Future Diaries", in* GENETIC SECRETS 67 (Mark Rothstein ed., 1997); Jacqueline Chin & Alastair Campbell, *What—if Anything—is Special about Genetic Privacy, in* GENETIC PRIVACY 225, 232 (Terry Kaan & Calvin Ho eds., 2013).

[167] Daniel Callahan, *Health Care Reform: Can a Communitarian Perspective be Salvaged?*, 32 THEOR. MED. BIOETH. 351, 360 (2011).

[168] *But see* 42 U.S.C. §1320e-1(c)(1); *id.* at §1320-1(e) (prohibiting use of quality adjusted life years to determine "what type of health care is cost effective or recommended.").

[169] *Cf.* DANIELS, *supra* note 44, at 24-25, 114-15 (explaining why tradeoffs can only be determined through legitimate processes).

2. Procedural Justice: Democratic Decisionmaking

While the principles of substantive justice provide broad principles of fairness in distribution at a general level, translating those principles into practice is not easy. Different substantive principles may point in different directions: calibrating burdens and benefits depending on welfare may require collecting information from wearables, but many may feel that this exceeds the scope of state power. Distributing benefits equitably may require collecting data from marginalized individuals, but these individuals often are worse-off as a group—we would nonetheless increase privacy burdens on them. These questions and contradictions must be resolved through discussion and debate, through a transparent and legitimate process that Smith's principles prescribe.[170]

I begin with outlining the existing system of individual informed consent and contrast it with the democratic consent model. I explain how both models rely on similar ethical premises, but argue that a democratic consent model addresses some of the legitimacy problems that individualized informed consent faces.

a) Contrasting models

Traditionally, research with identifiable data requires the subjects to consent on an individual basis after obtaining information about the research and potential benefits and risks. Information thus collected can be used only for the specified purpose.

The taxation model moves health information collection from an individual consent based, private, contractual model, to a public law, democratic consent model. Instead of relying on individuals to provide consent for their particular information, individuals will vote for representatives based on the policies they articulate with respect to information collection and its justifications as with taxes.[171] Representatives will enact policies that bind all citizens. Technological decisions and day-to-day management should still be left to the agency, but Congress and the

---

[170] DANIELS, *supra* note 44, at 27, like me, relies on process to address more particularistic aspects of distribution, and explains how procedure meshes with substance. "[F]air process…set[s] limits and priorities…[but] is limited by [substantive] obligations."

[171] Most individuals expect public authorities such as ethics committees or ministries to govern biobanks. Gottweis et al., *supra* note 193, at 436.

public should take a broader role in shaping ethical limits.[172] No longer should the health information collection project be relegated to technocratic backwaters, with politicians having merely a vague understanding of its benefits. [173] Regular democratic governance itself becomes the basis for collection and use, though we may devise special procedures for certain kinds of highly sensitive information.[174]

Some may dispute that I am suggesting anything new—after all, our elected representatives can, at any time, pass legislation to reflect the policy preferences of voters regarding health information collection. But as a practical and ethical matter, this is not the bargain we have struck. There is no civic discussion of what must be done in the health information context. Rather, we have delegated decisionmaking to administrators and IRBs. This departs from the procedures we actually utilize in the revenue taxation context where we delegate far less, debate principles in solidarity, cast votes based on our opinions regarding taxation, and delegate to the government only lower level considerations.

Finally, regular procedural and constitutional limits would apply to the process. Individuals can judicially challenge policies they believe offend higher order principles of constitutional democracy or administrative

---

[172] One objection is that we live in a state with imperfect democratic participation. Bioethicists may therefore argue that something more is required for decisions to be truly legitimate. For example, they may counsel the use of citizen juries, deliberation groups, etc., in developing appropriate policies. But the problem to which they point is not one isolated to bioethics contexts—it calls into question all coercive laws, many of which are far more invasive information collection. Answers lie more properly with theorists of democratic participation rather than with bioethicists. *See, e.g.,* Bruce Ackerman & James Fishkin, *Britain Should Deliberate Before it Votes on Europe*, HUFFINGTON POST (June 17, 2015), http://www.huffingtonpost.co.uk/bruce-ackerman/britain-should-deliberate_b_7607312.html.

[173] Some may object that the democratic decisionmaking model I portray is too idealized and simply does not exist in real life. Hence, I cannot realistically rely on it. Indeed, Daniels spends must time addressing participatory processes in his work. DANIELS, *supra* note 44, at 110-39. But ultimately, like me, Daniels settles on broad democratic processes. As he explains, other participatory mechanisms simply cannot "substitute for broader democratic processes." *Id.* at 130. Daniels goes on to describe how such processes can be assisted. For me, however, those are questions of democratic theory, best left to specialists in those areas, rather than of health law and bioethics, and Daniels himself recognizes the limits of what he can provide. *Id.* at 138.

[174] Some scholars have gestured slightly in this direction. *See, e.g.*, Sue Weldon, *Public Consent or Scientific Citizenship, in* GENETIC DATABASES, *supra* note 128, at 161, 162 (inquiring whether a "social equivalent of informed consent" may look like a form of "public consent" such as a "popular vote").

transparency. This is exactly what occurred in Iceland over a decade ago, when the Iceland Supreme Court invalidated a public-private partnership to collect and integrate detailed information regarding all Icelandic citizens.[175]

b) Analogous Ethical Grounding

This democratic consent model largely remains faithful to the key ethical principles that undergird individual informed consent. Both informed consent and democratic consent models derive from similar ethical principles of non-exploitation and respect. Relying on the work of Joel Feinberg and Alan Wertheimer,[176] Lynn Jansen has recently argued that the primary purpose of consent in research is to prevent exploitation of individuals by researchers.[177] In giving consent, the individual effectively participates in the enterprise. In a way, the project becomes a shared or joint enterprise of the patient along with the researcher. With true informed consent, the individual ends up sharing the goal of the researcher—and so is not used or exploited.[178]

Jansen's work suggests that consent is also important as a way to show respect to the individual.[179] By offering individuals the chance to participate in the decision, informed consent treats the individual as an entity that is able and deserving of making important decisions. It creates a forum in which individuals are able to express and exercise their preferences. Participation therefore is an important part of respecting and preserving the dignity of individuals.

---

[175] *See generally* David Winickoff, *A Bold Experiment: Iceland's Genomic Venture, in* ETHICS, LAW AND GOVERNANCE OF BIOBANKING 187, *supra* note 47, at 195 & n. 18. While others view Iceland's venture as an example of the dangers of letting "democratic will" through "presumed consent" triumph over regular "informed consent" mechanisms, *id.* at 199, to me, the story proves the opposite. It shows that when democratic will runs amuck, counter-majoritarian limits imposed on the majority can vindicate the rights of individuals.

[176] *See* Jansen, *supra* note 197, at 29. And popular debates in other countries have led to discussions. *See* Weldon, *supra* note 174, at 170 (noting the call for Parliament to take up the matter more thoroughly in the UK). Finally, note that I assume here that democratic participation is the best form of public participation, but other more participatory forms exist. Thus one ethicist conceptualizes a ladder of public participation, with different levels of participation. *See id.* at 175

[177] The anti-means principle and the non-exploitative principle are related but not identical. *See* Jansen, *supra* note 197.

[178] *Id.* at 30-31.

[179] This is a corollary of Jansen's non-exploitation argument, but important enough to be highlighted in its own right.

Although democratic decisionmaking occurs at a community-wide level, it comports well with Jansen's non-exploitation principle.[180] Individuals participate at a community level in the decision of how information should be used. Admittedly, individuals do not directly participate in the decision as Jansen envisages. Rather, the interaction is mediated by the community of which they are a part. It is the community, that, as a whole, adopts the goals of the research through the democratic process. In turn, each individual adopts the goals of the community of which she is a part long before the specific research encounter, as part of the social contract. Although some of those goals may not comport with the individual's preferences, the individual will have provided her input in shaping the goals of the community. The individual therefore indirectly participates in the research as part of a joint enterprise.

Further, the move to democratic decisionmaking also respects the dignity principle, but does so again, by treating the individual as part of a community. She does not act alone, unconnected to the rest of society, but rather, as a citizen, picking the goals towards which the research will be directed. To be sure, she does not have complete control over her own information, but in this context, she also has a say in the use of the information of others. She therefore is treated, along with others, as an autonomously functioning member of the community.

c) Solving problems

Even though it shares a similar ethical heritage as individual consent, democratic consent addresses key problems that plague the former.

First, individual consent in our current system it is a study in extremes. Data is either overly protected, or not protected at all. On one end, as prominent critics, including the Institutes of Medicine have noted, informed consent requirements which may require obtaining consent from the millions of individuals required for a particular study, can be practically

---

[180] The notion of presumed consent is admittedly not without problems., as is evinced by the 2003 ruling of the Iceland supreme court questioning how the legislature could impute assumed consent to Icelanders for collection of their health records. But a large portion of the reasoning had to do with the fact that deCode was benefiting and held a monopoly over this collection. The court explicitly denied that consent could be imputed for this kind of monopoly. *See generally* Kaye, *Abandoning Informed Consent, in* GENETIC DATABASES, *supra* note 128, at 117. It is unclear whether a governmental approach such as the one I advocate here would meet with a similar fate—and given our taxation system, it begs the question of whether it *should*.

impossible to meet.[181] This high level of protection, they argue, is overkill, given that the burdens of informational research are minimal and that patients often do not make choices with respect to other, perhaps more meaningful aspects of their treatment (e.g. location of hospital beds or sources the doctor uses to come to her conclusions).[182]

At the other extreme, data not covered by informed consent requirements is left completely unprotected, and strong regulation on par with states such as California has been stymied because of the necessities of research.[183] Numerous entities are not covered by HIPAA or Common Rule requirements, and therefore need not obtain informed consent.[184] Entities can collect data from non-medical sources, such as social media sites; the legitimacy of such collection and use, in the absence of agreements, is open to question. IRBs can (though often do not) release data without consent if certain conditions are met.[185] Covered entities may also engage in certain deidentification processes which allow them to share the information under HIPAA without obtaining business associate agreements, even though the information can be, in many cases, reidentified.[186] There is no evidence that this lack of protection tracks individual preferences.[187] By occupying

---

[181] *See, e.g.*, BEYOND, *supra* note 3, *passim*, and sources cited below.

[182] Cate notes that consent forms are hard to read, inadequate to motivate action, and discusses the absence or illusion of choice. Privacy protection is ultimately inadequate, but the red tape imposes burden on the individual and companies, Cate, *supra* note 14. Faden et al., *supra* note 69, at S53 opposes informed consent in the informational research context altogether, as such research imposes no real burden on the patient in exchange for great benefits. *See also* Cohen et al., *The Legal And Ethical Concerns That Arise From Using Complex Predictive Analytics In Health Care*, 33 HEALTH AFF. 1139, 1143 (2014) (noting that patients do not give consent and are unaware of sources physicians consult, bed allocation, etc. and questioning whether informational research is any different. Busby, *Blood Donation, in* GENETIC DATABASES, *supra* note 128, at 41.

[183] *Protecting Patient Privacy*, *supra* note 18, at 275 (discussing the term "health inflected" data).

[184] *See generally*, *Protecting Patient Privacy*, *supra* note 18.

[185] *Protecting Patient Privacy*, *supra* note 18, at 275; *see* W. Lipworth et al., *Consent in Crisis*, 36 INTERNAL MED. J. 124 (2006).

[186] *See, e.g.,* 45 C.F.R. § 46.101(b)(4) (exempting from the Common Rule "[r]esearch involving the collection or study of existing data, documents, records, pathological specimens, or diagnostic specimens, if these sources are publicly available or if the information is recorded by the investigator in such a manner that subjects cannot be identified, directly or through identifiers linked to the subjects"). The rule is partially altered in the NPRM, *supra* note 5, at 53,973.

[187] This data can be reidentified of course, Mark A. Rothstein, *Is Deidentification Sufficient to Protect Health Privacy in Research?*, 10 AM. J. BIOETHICS 3 (2010). But individuals are protective of even their deidentified data. *See* Peel, *supra* note 42.

extremes, the legitimacy of existing informed consent regulation (or non-regulation) invites calumny from all sources.

Second, there are collective action problems which further undermine the legitimacy of the process. First, although individuals get to veto the researcher's use of their particular information on their research project, they do not, *pace* Jansen, have a say on what projects the researcher may pursue. They cannot band together with other individuals to promote some projects and curtail others. Further, there is a free rider problem. Even if an individual believes in a project, there is an incentive to avoid providing data in the hope that the researcher can get data from other individuals.

Third, the ethical focus of informed consent is all about collection of data before the project; there has not been an equal focus on return of results after the project is done. Informed consent policies do not, generally, give individuals a right to mandate a certain quality of product, or even to get results back at all for their contribution. There is therefore attenuation on the back end of research projects.

Scholars have suggested two main alternatives to the traditional approach.[188] While addressing some of the problems I describe, they exacerbate others.

First, some address the problem of extremes by offering broad consent regimes as a middle ground. This model leaves essential aspects of informed consent intact. An individual simply consents to *all* possible research that could be done with her information. As scholars note, broad consent essentially involves delegating research decisions to the holder of the resource.[189] Some regimes are broader than others.[190] A recent Notice of Proposed Rulemaking would adopt this approach with respect to data and

---

[188] Some limited approaches, such as that of the IOM, propose adjusting the Privacy Rule to conform to the Common Rule, with less stringent anonymization requirements. However, it is unclear whether these changes will solve the problems outlined above to a significant degree. BEYOND, *supra* note 3, at 201.

[189] Studies have shown that while a majority of patients want to provide consent, a majority is also happy to delegate decisions to a Research Ethics Committee. Busby, *supra* note 182, at 44 ("[T]here was a sense in which donors, working at the limits of their expertise, chose to hand over trust, a choice with physical and emotional dimensions as well as the more evident congnitive elements"); Antonio Casado da Rocha, *Biobank Governance in Spain, in* ETHICS, LAW AND GOVERNANCE OF BIOBANKING 229 (Deborah Mascalazoni ed., 2015) (59% of Europeans opted for delegation).

[190] Sutrop & Simm, *supra* note 145, at 209-10 (listing forms of broad consent).

certain kinds of biospecimens in all federally funded research.[191] This approach is most prominent in the biobank context, and is utilized heavily in Europe.[192]

But this "middle ground" between extremes exacerbates the collective action issues underlying the lack of control over research. Concerns remain that broad consent deprives individuals of their veto power over inappropriate research. They may not truly understand the scope of the research that can be carried out with their information. Scholars have suggested that individuals be permitted to customize consent, for example, by allowing them to permit certain research uses, and forbid others.[193] Others would prohibit the holders of the resource from engaging in any controversial research. But in the most expansive broad consent regimes, individuals still do not have enough downstream control over their information.

Accordingly, scholars and policy makers have proposed allowing individuals to participate in decisions about what happens with their information. They propose various methods, including technological solutions, such as menu systems by which individuals can pick from a menu of research projects for which their data can be used.[194] Patients can also influence the choice of research projects,[195] through "citizen juries," or data

---

[191] NPRM, *supra* note 5, at 54,048-49. Currently, broader versions of blanket consent are impermissible. *See* 45 C.F.R. § 46,116.

[192] Rustam Al-Shahi et al., *Bias from Requiring Explicit Consent*, 331 BRITISH MED. J. 942 (2005) (noting that majority of the international bioethics community has supported the use of broad consent).

[193] Most people give broad consent for long term research on samples. Herbert Gottweis et al, *Biobanks and the Phantom Public*, 130 HUM. GENETICS 433, 436 (2011). Patients often do not take the time to understand the consent forms, and prefer instead to delegate decisions to institutions they trust. *Id.*

[194] Various consent management tools exist. Peel, *supra* note 42, provides an excellent overview.

[195] *See generally*, Haimes & Whong Barr, *supra* note 140; Hoffman & Podgursky, *Improving Health Care Outcomes*, ***supra*** note 5, at 430 (suggesting patient input as to what measures are important for them and their opinions on side effects, pain, recovery time, personal appearance).

cooperatives that provide input or vote on how information can be used.[196] For some ethicists, such participation vitiates the need for consent.[197]

But even these solutions do not truly guarantee stakeholders actual participation in the decisionmaking process. The regimes are often developed unilaterally by private entities, without open decisionmaking or accountability to the participants. The British biobank's "group consultations" have been criticized on precisely this ground—how, authors have asked, will the biobank assess opinions, whose opinions are relevant, what weight will they play, and will they affect outcomes?[198]

Democratic consent tweaks the broad consent and participatory solutions. In the democratic model, the social contract itself is a form of broad consent. Individuals bind themselves in advance to providing for whatever outcome society as a whole or the government that represents it, decides is appropriate. And the democratic model guarantees participation in the form of elections, in ways far more robust than the participatory mechanisms suggested thus far.

Democratic consent addresses the key problems of individualized consent. First, democratic participation does not indulge in extremes. Data is protected per the wishes of the collective with respect to all data, not per the wishes of each individual with respect to her own data. Although the contractual or ethical bond between the patient and downstream recipients of the information may never exist or become severed due to contractual attrition,[199] the government's special relationship with its citizens will always remain. When the government accesses data within constitutional limits pursuant to democratically enacted laws, its access is legitimate.[200]

---

[196] Donald Willison, *Privacy and the Secondary Use of Data for Health Research*, 8 J. HEALTH SERVICES RES. & POL'Y S1, 17 proposes citizens jury to get degrees of consent right. There is an Emphasis in the literature on donors and community members actively engaged in process of research governance, *See, e.g.*, Kozlakidis et al., *supra* note 21, at 11. For coops, see Richard Tutton, *Constructing Participation in Genetic Databases*, 32 SCI. TECH. & HUM. VALUES 172 (2007).

[197] *See generally* Lynn A. Jansen, *The Ethics of Altruism in Clinical Research*, 39 HASTINGS CTR. REP. 26, 29 (2009); *see also* Christensen, *supra* note 119, at 110 (arguing that individuals would help if they could also direct the course of research instead of merely consenting.)

[198] Alan Petersen, *Securing our Genetic Health*, 27 SOC. HEALTH & ILLNESS 271, 283 (2005) notes that the "Consulation model" for the UK biobank is questionable. He worries that the consultants never explained how opinions would be taken into account, and pointed to problems with the assessment of opinions. Thus questions arise with respect to the sufficiency of process. *See also* Hoffman & Podgursky, *supra* note 3, at 129.

[199] See Kaye, GENETIC DATABASES, *supra* note 180, at 107.

[200] *See New Infrastructural Model*, *supra* note 33, at 631.

And it always remains subject to suit under robust privacy laws that apply to government-held or accessed data,[201] and to electoral influence, if it misuses the data. Finally, since the government can ensure access to research data, more robust regulations that impose privacy protections by all private researchers may become practically and politically feasible.

Second, the approach addresses collective action problems. Unlike in Jansen's account, when democratic decisionmaking occurs, we do not assign separate roles to patients and researchers, with the former providing the input into research and the latter directing research. Rather, the entire community comes together to direct research goals and secure the inputs for that research.[202] Thus, democratic participation can bring research priorities in line with existing preferences.[203] If patients experience the indignity and inconvenience of illness in different ways than health care providers anticipate, we may see a shift in research priorities to more accurately reflect what patients care about.[204] As patients gain more benefits from the system, their level of compliance with information collection mandates will potentially improve as well.

On a related point, Jansen expresses concerns that under an individual informed consent model, individuals may over-participate in research because they may show too much altruism, or possibly even self-hatred. Under a community-decisionmaking model self-hating or over-altruistic individuals will always be a minority of the vote; the majority would prevent over-exploitation of such individuals.

Finally, unlike private entities, because of their special relationship, the government owes its citizens a duty to provide results from the system, at a community wide or individual level. These results must meet minimum

---

[201] *See, e.g.*, The Privacy Act of 1974, 5 U.S.C. § 552a.

[202] Shawn Harmon, *Solidarirty: A (New) Ethic for Global Health Policy*, 14 HEALTH CARE ANALYSIS 215 (2006). There is some suggestion as well that because the different shape of research means that stakes of individual decisions are so high—the decision does not primarily involve whether the individual will survive, but whether humanity will benefit—it is not appropriate to place the whole decision on him. *See* Klaus Hoeyer, *Ambiguous Gifts: Public Anxiety, Informed Consent, and Biobanks, in* GENETIC DATABASES 97, *supra* note 128, at 102, 111.

[203] Richard R. Sharp & Mark Yarborough, *Additional Thoughts on Rethinking Research Ethics*, 5 AM. J. BIOETHICS 40 (2005). *See also* Amy L. McGuire et al., *Ethical and Practical Challenges of Sharing Data from Genome-wide Association Studies*, 21 GENOME RES. 1001 (2011) (explaining the need for public trust, open discussion, deliberative democracy, for increasing trust; Rebekah McWhirter et al., *Community Engagement for Big Epideiology*, 4 J. PERS. MED. 459, 461 (2014) (Tasmania) (same)

[204] *See generally* Weldon, *supra* note 174.

quality standards. Citizens can debate the exact level of benefit they wish to obtain and enact that into law. The next Part describes the steps the government must take to produce these goods.

3. Efficiency: Ease and Minimizing Waste

Smith's final two principles seek efficiency—information should be collected with ease and minimum waste. The tax system has settled on a process of centralized government collection, collation, and dispensation of revenue. Private entities, such as employers or investment companies, will sometimes assist the government by withholding taxes.[205]

As other scholars have described in some detail,[206] the current system of health information collection and secondary research is largely driven by private entities, though the government is increasingly taking on a leadership role. Under private control, problems plague the secondary research process at numerous stages including (i) pooling the information (ii) accessing the information for research (iii) maintaining information quality, (iv) determining research priorities, and (v) allocating the goods produced through the research. Many of these problems result from a lack of coordination and anti-competitive tendencies among private firms.[207]

As a matter of procedural justice, I have argued that the government should take the lead on information collection. Government collection will also help address the inefficiencies that beset health information collection.

---

[205] *Cf. supra* note 85.

[206] *See, e.g.*, sources in *supra* notes 18, 33.

[207] The literature has generally uniformly assumed (or advocated) that private entities collect identified data and have final say over the direction of the research, and that the central government defers to state and local mandates. AM HEALTH MGMT ASSN, DEVELOPMENT OF A NAT'L HEALTH DATA STEWARDSHIP ENTITY, RESPONSE TO REQUEST FOR INFORMATION 17 (Aug. 3, 2007) ("There should be no central repository of aggregate data, whether at the national or regional level"). A single data repository for aggregating and reporting quality data could fail to meet user needs, increase the risk of large scale privacy violations and undermine public trust. Rodwin, *supra* note 210, at 594. Rodwin and, following him. *Finding a Cure*, *supra* note 33, pushes for public ownership, but only of deidentified data. This is useless for many of the research projects this Article seeks to promote. Industry in particular has opposed government monopolies. Jeffrey R. Gulcher & Kári Stefánsson, deCODE genetics, *The Icelandic Healthcare Database and Informed Consent*, 342 NEW ENG. J. MED. 1830 (2000) (executives of a U.S. based company explaining, "It is probably better for a private company to hold this information than for the state to do so, since governments can violate the privacy of individuals to advance the interests of society as a whole. Moreover, if a health care data base managed by a private company violates privacy, the company can be closed down.").

Government leadership can take the form of centralized collection of data to simply robust regulation of private efforts to create a centrally accessible distributed network. More research is required to identify the best approach.

Finally, I anticipate that private entities would be allowed to continue to use the data they collect. Citizens also may elect to engage in what scholars have referred to as citizen science—creating private cooperative research arrangements by pooling their own, or accessing previously collected data.[208] Government curated data accessed via all these private entities may be superior to the data than any one entity or group of entities can provide. But these smaller entities can continue to carry out innovative cutting-edge work using their proprietary data that can serve as models for larger government projects.

a) Correcting Inefficiencies

Problems beset each stage of the collection and deployment of information. Each of these problems result from coordination problems and anticompetitive tendencies among firms. The government is best positioned to address these concerns.

First, pooling data is necessary as "even the larger payers—apart from government—do not possess the critical mass necessary" to carry out meaningful research.[209] But pooling can prove challenging because of interoperability problems, or possible antitrust problems, among others.[210] Further, there will be high transaction costs in overseeing pooling by private entities.[211]

Second, even pooled data may be in- or unevenly accessible. IRBs are notoriously cautious and inconsistent in providing waivers to access identified data, without which researchers may require millions of patients to sign informed consent documents.[212] Multi-institutional projects (involving pooled

---

[208] Sharona Hoffman, *Citizen Science: The Law and Ethics of Public Access to Medical Big Data*, 30 BERKELEY TECH. L.J. (forthcoming 2015)

[209] BASIC STAPLE, *supra* note 210, at 251.

[210] Marc Rodwin, *Patient Data: Property, Privacy & the Public Interest*, 36 AM. J.L. & MED. 586, 501 (2010).

[211] *Id.* at 613.

[212] *See supra* note 50. The Common Rule NPRM, *supra*, note 5, would greatly loosen these restrictions, but to the extent the research I describe relies heavily on clinical entities covered by HIPAA, restrictions will remain until HIPAA regulations are altered. *See id.* at 53,944 (recognizing that HIPAA restrictions on data sharing will remain in place, and that those restrictions are stricter than that of the existing Common Rule). But the proposed relaxing of the Common Rule requirements is premised on the promise that HIPAA will continue to

data) are even harder as many institutions require their own IRB—each with its idiosyncratic application of the relevant standards—to decide whether to permit access to data.[213] Further, over time, data may be lost as biobanks or companies fold, merge, change research focus, or decide that they no longer want to expend the resources to store the information or biological samples.[214] And companies not subject to the Common Rule or HIPAA often limit access to proprietary information for competitive reasons, to stifle research, and due to fear of liability.[215]

Third, even if the data is pooled and is accessible for research, there may be quality problems. The problems may arise when the data is initially transmitted if the entity from which the information originates uses substandard technology.[216] Mistakes may arise if the entities use different platforms. Further, pooling data held by different organizations maintained

---

limit data use, so any change in HIPAA does not seem imminent. *See id.* at 53,953. Further, the Common Rule will remain a floor; states and institutions can demand more, *id.* at 54,046, and, unlike the current regime, patients will be able to completely block the use of their information even if a IRB deems the information use harmless. *Id.* at 53,976.

[213] BEYOND, *supra* note 3, at 224-26.

[214] R.J. Cadigan et al, *Neglected Ethical Issues In Biobank Management*, 9 LIFE SCI & SOC'Y & POL'Y 1 (2013) (noting that this is a neglected ethical issues in biobank management).

[215] In the status quo, private companies and doctors restrict access to information due to concerns involving competition and legal liability. Laakmann, *supra* note 15, at 16, 21. Frank Pasquale, *Grand Bargains for Big Data*, 72 MD. L. REV. 682, 711 (2013). BASIC STAPLE, *supra* note 210, at 194. A study of 500 HIPAA privacy rule cases showed that the most common cause for non-release of health data by providers to the patients themselves was fear of liability. Sara Rosenbaum, *Data Governance and Stewardship*, 45 HEALTH SERVICES RES.1442, 1443 (2010). BASIC STAPLE, *supra* note 210, at 124, 190. But other reasons can exist. Competing health care organizations that treat overlapping patient populations in a community may be reluctant to share relevant data, typically because each organization fears that others could use its data for competitive advantage. Joachim Roski et al., *Creating Value In Health Care Through Big Data: Opportunities And Policy Implications*, 33 HEALTH AFF. 1115 (2014). Joshua Rolnick, *Aggregate Health Data in the U.S.*, 19 HEALTH INFORMATICS J. 137, 145 (2013) lists the astronomical costs of medical database searches for informational research. The case of Iceland's database, over which U.S. commercial firm DeCODe had a monopoly, is instructive. As scholars noted, after the monopoly was invalidated by the nation's supreme court, it seems unlikely that the citizens of that state would willingly make a large gift of their DNA to a for profit U.S. corporation. G.J. Annas, *Rules for Research on Human Genetic Variation — Lessons from Iceland*, 342 NEW ENG. J. MED. 1830, 1831 (2000). There are some very limited signs of changing trends, such as funders requiring data sharing plans in grant applications. Patricia Kosseim et al., *Building a Data Sharing Model for Global Genomic Research*, 15 GENOME BIOLOGY 430 (2014).

[216] *See* Matteo Ferrari, *Conveying Information, Generating Trust, The Role of Certification*, in COMPARATIVE ISSUES IN THE GOVERNANCE OF BIOBANKS 281, 282 (Giovanni Pascuzzi et al. eds., 2013).

in separate repositories may result in record duplication.[217] Ultimately, inaccurate data "might lead to patients being harmed rather than helped."[218]

Finally, assuming that the data is properly pooled, accessible, and of high quality, there is evidence to suggest that private entities will focus only on particular causes and underutilize information.[219] Pharmaceutical companies have overproduced me-too therapies for lucrative markets, and under-produced more essential therapies for less lucrative ventures, such as Ebola treatments or antimicrobials.[220] Non-profits in turn may focus only on specific projects. Even if these entities painstakingly collect information, after their project they may discard the data, or delete elements not relevant to their work.[221] And private entities in the past have carried out research in problematic ways, including by ghost authoring research papers or dropping patients (in this case, records) that demonstrate bad results.[222]

Each of these problems can compound each other—for example, in a system where companies may only choose to collect the data of, research into, and disseminate benefits to, groups that are economically beneficial to them, health disparities may well intensify exponentially because of the pressures at each step of the process.

Of these problems, scholars have primarily focused on the question of coordination. They have suggested consolidation of data pools to obtain better research quality and using intermediaries to address fragmentation or coordination problems—but almost none have advocated for government control. The Institutes of Medicine recommended a "trusted intermediary that could link data from different sources and then provide more complete and useful deidentified datasets."[223] Inspired by Swiss approaches, Alex Pentland and coauthors suggest a data cooperative, which is run by a one

---

[217] BASIC STAPLE, *supra* note 210, at 254.

[218] *Id.*

[219] *Cf.* Cadigan et al., *supra* note 214, at 5 (arguing that biobank resources may similarly be underutilized).

[220] *See, e.g.*, Rennie, *supra* note 80, at 42-45; Corrigan, *Informed Consent, in* GENETIC DATABASES, *supra* note 128, at 78, 79 ("what distinguishes [private] data collection from other forms of regional or national databases is its global dimension audits lack of public knowledge of or accountability for research carried out on commercial databases.").

[221] This is not a relevant concern when the holder is a generalized databank like that of the Kaiser system for example.

[222] Efthimios Parasidis, *Patients over Politics*, 2011 WIS. L. REV. 929, 975-76 (2011). An analogous move is taking place in the observational research context. For example, companies are focusing only on well to do communities to avoid bad outcomes. INTEGRATING, *supra* note 4, at 138.

[223] BEYOND, *supra* note 3, at 191.

member one vote principle. The co-op would sell data and invest the revenue. Although this is a private system, Pentland refers to this as "true citizen empowerment."[224] Barbara Evans recommends a system of trusted intermediaries, overseen by a regulatory body supported by fees.[225] Fred Cate calls for centralized data collection centers.[226] Rodwin acknowledges that the expectation was that non-federal Health Information Exchanges (and before them, Regional Health Information Organizations) would consolidate in a "nation-scale system" to aid in data pooling.[227] Other advocacy groups recommend a temporary collection and honest broker system.[228] And several health systems have begun consolidating IRBs.[229]

These accounts lay the basis for what the taxation analogy points to—centralized government coordination.[230] Government can pool information more easily than other entities. It also is less subject to the conflicts of interest in maintaining information access, and setting research priorities and distributing benefits than private entities. Further, it has the coercive authority to maintain information quality. One can imagine, for example, the government mandating that all individuals provide information, and perhaps, even setting an annual data quality day by which individuals must check for errors in their EHRs, much like the tax filing deadline.

Administrative processes could be modeled to some degree based on the tax analogy. So far, each sub-agency has collected information for agency specific projects: FDA for example collects prescription data, CMS Medicare data, and so on. The Office of the National Coordinator for Health Information Technology (ONC) is involved in only a supporting capacity. This

---

[224] PENTLAND ET AL., *supra* note 22, at 27.

[225] Evans, *supra* note 2, 112-113 (HITECH act is a cost based service, that refunds practitioners because of the costs they put into the project).

[226] Cate, *supra* note 20, at 1798.

[227] Rodwin, *supra* note 210, at 595.

[228] ACADEMY HEALTH, LEGAL AND POLICY CHALLENGES TO SECONDARY USES OF INFORMATION FROM ELECTRONIC CLINICAL HEALTH RECORDS, *available at* http://www.academyhealth.org/files/publications/HIT4AKLegalandPolicy.pdf.

[229] INTEGRATING, *supra* note 4, at 25 (New York); 79, 82 (Cincinnati).

[230] Don Detmer is the sole academic who has firmly aedvocated for federal consolidation, no matter the health data. He notes that historically, government leadership in this area has always been envisaged. Don Detmer, *Building the National Health Information Infrastructure for Personal Health, Health Care Services, Public Health and Research*, 3 BMC MED. INFORMATICS & DEC. MAKING 1, 4 (2003). *See also* Rosenbaum, *supra* note 215, at 1452 (advocating for a required data federal submission category in passing).

may, in some cases, lead to waste and duplication—one can imagine multiple agencies accessing data from the same doctors regarding the same patients.

The tax analogy points to the virtues of centralizing collection processes in the hands of one agency. ONC should be in charge of collecting or maintain access to health information. Entities in other agencies can then be given access to the information as required. Exact details must await further research.[231]

b) Considerations Against Government Leadership

Two concerns counsel against government leadership. First, some suggest that individuals are more wary of government than industry data collection. But research points in both directions. Research finds that individuals are content with letting data be collected for research, and specifically, government research, but are far more wary of commercial use of the data.[232] Other studies have found distrust of government collection (where it was unclear whether the *purposes* of government collection were elucidated.)[233] There is, for example, a great likelihood that individuals will oppose law enforcement access to the data.[234] This argument against government collection is, for now, not compelling.

More damningly, however, are security considerations—with all the data in one place, a single breach could be devastating. Nonetheless, once various security measures are put in place, it is far from clear that private

---

[231] Models abound. *See, e.g.,* Petersen, *supra* note 198, at 277 (discussing ideal relationships between industry and government). HHS divisions are already trying to determine the division of responsibility between agencies. FDA, FDASIA HEALTH IT REPORT 26 (April 2014), *available at* http://www.fda.gov/downloads/AboutFDA/CentersOffices/OfficeofMedicalProductsandTobacco/CDRH/CDRHReports/UCM391521.pdf. While I have policy recommendations on this front, those are best saved for a different article.

[232] BASIC STAPLE, *supra* note 210, at 236; Studies show that individuals are sympathetic to data being used for medical research, but are wary of the data being transferred to the commercial sector, even if it is anonymized. Claudia Slegers et al., *Why Do People Participate in Epidemiological Research*, 12 J. BIOETHICAL INQUIRY 227 (2015). Confidence in industry has been undermined by undermined in recent years by financial incentives to undertreat or overtreat patients, conflicts of interest involving pharmaceutical companies, and limits on access to physician services attributable to managed care; and commercial use of information. Julian Sheather, *Patient Confidentiality in a Time of care.data*, 347 BRITISH MED. J. f7042 (2013).

[233] Rothstein, *supra* note 187; L.J. Melton III, *The Threat to Medical-Records Research* , 337 NEW ENG. J. MED. 1466 (1997).

[234] Petersen, *supra* note 198, at 278.

entities have any advantages over the government when it comes to data security. Indeed, private firms have experienced breaches that have been comparable or worse than the government's.[235]

First, technological approaches can be used. Data can be stored in multiple locations, and can be linked for specific searches (though designers may also need to consider latency costs and duplication).[236] Further, records can be linked solely by identifying biomarkers rather than by names or social

---

[235] For an interactive infographic, *see* Sisi Wei & Charles Ornstein, *Over 1,100 Health Data Breaches, But Few Fines*, ProPublica (Feb. 27, 2015), *available at* https://projects.propublica.org/graphics/healthcare-data-breaches (documenting 1,199 large-scale data breaches since October 2009) (citing U.S. Dep't of Health & Human Servs., Office for Civil Rights, *Breaches Affecting 500 or More Individuals*, U.S. DEP'T OF HEALTH & HUMAN SERVS., https://ocrportal.hhs.gov/ocr/breach/breach_report.jsf). Government entities have also suffered breaches, but at declining rates. *See* Michael Froomkin, *Government Data Breaches*, 24 BERKELEY TECH. L.J. 1019, 1026 n. 35 (2009).

[236] Diaamond describes the process as follows:

> First, instead of collecting all of the detailed data, a distributed model would collect only summarized data (counts, numerators and denominators, or key results) and limit the data collection to the minimum needed to answer the research question. Second, personally identified information is held only at the source, and data are cleaned and analyzed in a common way at the source before being sent in a standardized format. Third, digitized data could be made accessible for analysis by different authorized entities across a network without requiring that each entity obtain a local copy of all of the detailed underlying data.

See generally, Diamond et al.,, *supra* note 42. Attempts to collect data for each population health initiative place a huge burden on data providers, who must field many requests for their data and report them repeatedly in many ways to different repositories. There is also the issue of privacy and security. As multiple or redundant large datasets *See also id.* Evans, *supra* note 2, at 99-100. Diamond, however, warns of a potential latency affect . This may be ok for long term tracking systems like FDA's mini-Sentinel system, *see* Mini-Sentinel, *supra* note 35, but cannot work for real time public health interventions. As Diamond puts it, long-latency systems "look[] less like surveillance and more like after-the-fact accounting." Diamond et al.,, *supra* note 42, at 454. There may also be duplication related issues without centralized storage. *See* BASIC STAPLE, *supra* note 210, at 114-15, 150-51. Data lake approaches may offer a new alternative. PWC. *Data Lakes and the Promise of Unsiloed Data*, PWC, http://www.pwc.com/us/en/technology-forecast/2014/cloud-computing/features/data-lakes.jhtml. An alternative is a system that collects core data elements in a central location with the remaining elements distributed, as envisaged in the PMI initiative. PMI Report, *supra* note 16, at 37.

security numbers, making them less readily identifiable. "Technological due process" may demand additional security measures.[237]

Second, the literature suggests that most HIPAA violations are caused, not through malicious intent, but by unawareness of applicable protocol. Ensuring that employees across numerous organizations are educated about privacy protections is harder than ensuring that employees in one entity are properly trained. Private firms "experience frequent staff turnover, which results in a continual challenge of adequately educating" the staff.[238] Training government workers in a centralized location may be far easier than ensuring that disparate companies with workforces with different skill and pay-levels and high turnover rates meet training standards.

Finally, centralization helps us watch the watchers.[239] As Nicolas Terry observes, "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. Indeed, the truth lies in the opposite because the big data business model is selling information about their data subjects."[240] Others argue that private companies—such as accountable care organizations and insurance companies—may breach data privacy to save costs by discriminating based on health status, or use information in inappropriate ways[241] Companies may provide access for benevolent but misguided reasons. Many, for example, try to maintain open access to their data to promote research, which increases the possibility of security breach. A centralized government point of access can help us ensure that "only bona fide researchers can obtain access…to preserve privacy and confidentiality."[242]

CONCLUSION

---

[237] Pasquale, *supra* note 215, at 726 n. 207 (suggesting audit trails); *See also* Danielle Citron, *Technological Due Process*, 85 WASH. U. L. REV. 1249 (2008).

[238] BASIC STAPLE, *supra* note 210, at 195.

[239] *See* JUVENAL, 6 SATIRE lines 347–8, THE LATIN LIBRARY, *available at* http://www.thelatinlibrary.com/juvenal/6.shtml ("Quis custodiet ipsos custodies.").

[240] *Protecting Patient Privacy*, *supra* note 18, at 5

[241] Hoffman & Podgurski, *Improving Health Care Outcomes, supra* note 5, at 434 (noting the fear that that learning health system providers will furnish physician's use of the system to health insurers who in turn might refuse to pay for treatments where physician deviated from the system's recommendations.). Further, insurance companies growingly discriminate based on health status. Pasquale, *supra* note 215, at 725 n. 204.

[242] Donna Gitter, *The Challenge of Achieving Open Source sharing of Biobank Data, Conveying Information, Generatiung Trust, The Role of Certification*, in COMPARATIVE ISSUES IN THE GOVERNANCE OF BIOBANKS, *supra* note 216, at 180, 185.

I have laid out a system of principles to guide policy in the new state-organized health information infrastructure. Much like private income, most health-pertinent information will probably not be part of this infrastructure, and remain in private hands. But that information designated as necessary for basic medical progress should be collected as part of a single system coordinated by a single government agency. This collection imposes various privacy burdens and risks, which have been disproportionately shouldered by members of society that are less well off, such as welfare recipients, and produces various benefits that have been denied to marginalized individuals. Burdens should be distributed equitably across society in ways that minimize the encumbrances on the least well-off. Benefits should be distributed to alleviate hardship, and translated into products and services that are made accessible to all individuals who contribute to the health information system.

To achieve these goals in a legitimate and efficient fashion, the government should take the lead in collection. This does not take from the companies that are engaging in data mining right now the ability to use their data in various ways. It simply mandates that the government also coordinate access to the data. Such mandates already exist in isolated contexts; federal efforts have never been subject to serious legal challenge.[243]

Implementation will take place in various stages. As I note above, these principles require shifting from a pragmatic demand-side research-question based approach to collecting data, to an ethical supply-side approach. This will alter which channels of data collection we focus upon, and influence the kinds of projects entities such as PCORI and PCORNet support through their grantmaking power. Similarly, meaningful use goals should be shifted away from CMS focused networks to the health system more generally. Networks and entities that are able to implement equitable risk-shifting should receive priority. The relevant stakeholders should work with these entities to identify groups that have the greatest medical need and understand how individuals experience information burdens.[244] In the long run, the vision documents for the National Health Information Infrastructure should incorporate these principles to ensure equitable distribution of privacy risk.

---

[243] Gobeille v. Liberty Mut. Ins. Co., 136 S. Ct. 936 (2016) held that *states* were preempted from imposing information collection requirements for all payer claims databases.

[244] Methods of calculating medical need vary. *See* DANIELS, *supra* note 44, at 230. This question will implicate numerous questions, of course, including how to define "medical" need, issues of jurisdiction, and other considerations.

But beyond these individual agency efforts, this Article calls for a sea change in the way lawmakers and citizens engage with health information collection. The principles and policy examples I offer are important not just for their substance and the policy they will guide, but also for the role they play in the new health information collection debate. They offer purchase, a starting point from which to start policy discussions regarding the new health information collection. Although individuals are often averse to taxation, its well-established place in popular culture means that it will give individuals the tools to conceptualize what is at stake, and stimulate and support discussions. Whether or not there is broad agreement as to these claims, they will now serve as a basis for future policy debates, as a foil for opponents, and as a foundation for their supporters.

Recognizing that health information collection practices implicate community wellbeing, equality, and justice, individuals should call on elected candidates to discuss them in their platforms. Legislators in turn, should participate in and encourage debate, solicit feedback from constituents, and more closely oversee agency practices. HHS agencies, similarly, should undertake innovative outreach efforts to engage individuals beyond those involved in the health collection industry.

The outcome will be piecemeal and pluralistic, albeit unified within a single system. As conversations progress, different theories of health information collection will be tested. Animated by different ideologies and approaches towards benefits, redistribution, and solidarity, individuals will push for different approaches; legislative compromises will often result in a morass of policies, each of which reflect different understandings of health information collection. The process, will, to be sure, stall or delay some collection efforts. But overall, the result will be a system that is more legitimate and transparent—after all, any delays will be the result of democratic processes.

My belief is that this engagement will ultimately redound to the benefit of the health collection enterprise.[245] Scholarship in the tax context has shown that as individuals internalize the values of taxation, they begin to comply with its dictates voluntarily. Their own identity and relationship with government around revenue collection is transformed—they move from

---

[245] Existing information suggests that individuals are in fact very supportive of the research enterprise even over privacy norms, when their autonomy is respected. *See* NPRM, *supra* note 5, at 53,944.

individuals arrayed against an adverse state trying to take their money to citizens participating in a joint enterprise mediated by government.[246]

Similarly, in the health information context, as individuals engage in the health information debate in more depth using familiar principles from the taxation context, we may see a shift understandings so that individuals see themselves, not just as patients, but as citizens, participating in a collective project in which they will have a stake.[247] This in turn may even make individuals fear privacy harms less as they perceive themselves as controlling outcomes.[248] But whatever the level we ultimately chose as a society to contribute health information, citizen participation is essential to maintain the viability of the system. Ultimately, the pursuit of the debate itself reifies health information collection as a site for enacting visions of citizenship, of relationships between the individual and the collective that define the obligations that individual and society can place on each other.

---

[246] Assaf Likhovski, *"Training in Citizenship": Tax Compliance and Modernity*, 32 LAW & SOC. INQUIRY 665 (2007).

[247] Citing cogent ethical reasons for revenue collection appears to improve compliance. *See* Haimes & Whong-Barr, *supra* note 128, at 69-70 (discussing various justifications and suggesting that with peer pressure and self-identification; individuals feel a "generalized cultural pressure or imperative to donate"). In one study, individuals noted that even in *biosample* collection, a majority will give up the data IF they are consulted first. *See also* B.A. Tarini et al., *Not Without my Permission*, 13 PUB. HEALTH GENOMICS 125 (2009).

[248] *Cf.* Hoeyer, *supra* note 202, at 106 ("When people feel that they are in control of tehcnologies they are much more likely to deem them benevolent.").