MUCH ADO ABOUT DATA OWNERSHIP

forthcoming in HARVARD JOURNAL OF LAW & TECHNOLOGY, Vol. 25 (Fall 2011)

Barbara J. Evans[*]

CONTENTS

Suggested citation: Barbara J. Evans, *Much Ado About Data Ownership*, 25 HARVARD JOURNAL OF LAW & TECHNOLOGY (*forthcoming* 2011), available at: http://ssrn.com/abstract=1857986

---

[*] Associate Professor; Co-director, Health Law & Policy Institute, University of Houston Law Center, bjevans@central.uh.edu. J.D., Yale Law School; Ph.D., Stanford University; Post-doctoral Fellow, The University of Texas M.D. Anderson Cancer Center. This research has been supported by the Greenwall Foundation and by the University of Houston Law Foundation.

I. INTRODUCTION

Who owns the data held in electronic health information systems is a question of nominal importance that threatens to distract from more pressing work that needs to be done to protect privacy while realizing the public health benefits of interoperable health data networks.

There is wide dissatisfaction with the HIPAA[1] Privacy Rule[2] and the Common Rule,[3] two key federal regulations affecting health information privacy. These regulations are criticized both for hindering access to health data[4] and for allowing too much data access.[5] In response, there have been calls to clarify data ownership.[6] A diverse array of proposals has emerged. One hears privacy advocates calling for patient ownership of data as a way to enhance patient privacy,[7] even as some scholars suggest it would make data more widely available for research.[8] Still others call for public (governmental) ownership to enhance researchers' access to data.[9] The one common theme is that property rights in data are important and that clarifying them should be high on the legislative agenda.[10] Ominously, this view is starting to infect

---

1 Health Insurance Portability and Accountability Act of 1996 (HIPAA), Pub. L. No. 104-191, 110 Stat. 1936 (codified as amended in scattered sections of 18, 26, 29, 42 U.S.C.).

2 45 C.F.R. pts. 160, 164.

3 Federal Policy for the Protection of Human Subjects of Biomedical and Behavioral Research ("Common Rule") 45 C.F.R. § 46.101–.124.

4 COMM. ON HEALTH RESEARCH AND THE PRIVACY OF HEALTH INFORMATION, INST. OF MED., BEYOND THE HIPAA PRIVACY RULE: ENHANCING PRIVACY, IMPROVING HEALTH THROUGH RESEARCH 66 (Sharyl Nass, Laura A. Levit, and Lawrence O. Gostin, eds., 2009), available at http://www.nap.edu/catalog/12458.html. [hereinafter, "IOM, PRIVACY REPORT"]; *see also* William Burman & Robert Daum, *Grinding to a Halt: The Effects of the Increasing Regulatory Burden on Research and Quality Improvement Efforts*, 49 CLINICAL INFECTIOUS DISEASES 328 (August 1, 2009) (arguing that" the application of the Health Insurance Portability and Accountability Act to research has overburdened institutional review boards (IRBs), confused prospective research participants, and slowed research and increased its cost.").

5 IOM, PRIVACY REPORT, *supra* note 4, at 66 (noting that the HIPAA Privacy Rule has not eliminated concerns of the public, which is "deeply concerned about the privacy and security of personal health information," and reporting, "In some surveys, the majority of respondents were not comfortable with their health information being provided for research without notice and express consent.")

6 *See, e.g.*, Mark A. Hall, *Property, Privacy, and the Pursuit of Interconnected Electronic Medical Records*, 95 IOWA L. REV. 631, 651 (2010) (arguing, "If patients were given ownership of their complete medical treatment and health records, they could license to compilers their rights to that information in a propertized form that could be more fully developed and commercialized."); Marc A. Rodwin, *Patient Data: Property, Privacy & the Public Interest*, 36 AM. J. LAW & MED. 586 (2010) (arguing for public ownership of de-identified patient data).

7 *See, e.g.*, Leslie A. Saxon, *Owning Your Health Information: An Inalienable Right*, THE HUFFINGTON POST (Oct. 7, 2009).

8 Hall, *supra* note 6, at 651; *see also* Mark A. Hall & Kevin A. Schulman, *Ownership of Medical Information*, 301 JAMA 1282, 1283-84 (2009) (discussing advantages of patient-controlled longitudinal health records and suggesting that one way to foster the development of such records would be to grant patients a right sell access to their records that is superior to the rights of entities that currently hold patients' data).

9 Rodwin, *supra* note 6, at 586; *see also* Marc A. Rodwin, *The Case for Public Ownership of Patient Data*, 302 JAMA 86 (2009) (arguing for governmental ownership of de-identified patient data).

10 *See, e.g.*, Hall, *supra* note 6, at 637 (stating, "The law's uncertainty over ownership and control of medical information is widely regarded as a major barrier to effective networking of EMRs [electronic medical records], and policy analysts consider the legal status of medical information to be a critical question at or near the top of issues needing resolution.") and at 631 (claiming, "How this issue is resolved can determine how or whether massive anticipated developments in electronic health records will take shape."); Rodwin, *supra* note 6, at 586 (claiming, "How the law defines ownership of patient data will shape whether its benefits can be developed and also affects patient confidentiality.").

policymakers,[11] raising a real risk that what began as abstract scholarly debate may end in ill-advised legislation.

Who owns the health data in administrative[12] and clinical databases is a matter of state law, in the absence of preemptive federal legislation clarifying data ownership.[13] A few state courts have held, in certain contexts, that health information belongs to the patient it describes.[14] Several states recognize patient property rights in at least one subcategory of health information (genetic information).[15] In the majority of states, ownership of health data is simply ill-defined.[16] A database operator has a legal interest in the data held within its system but this is not generally regarded as ownership.[17] Better characterizations of the database operator's role include "data steward,"[18] "data-holder,"[19] "data source,"[20] or simply "healthcare data environment."[21]

The urge to "propertize" health data needs to be weighed skeptically and with a clear understanding of how property rights actually work. If pursued, the propertization of health data may disappoint many of its proponents because of a surprising truth: the framework of patient

---

[11] *See, e.g.*, HOUSE COMMITTEE ON PUBLIC HEALTH, INTERIM REPORT TO THE 82ND TEXAS LEGISLATURE 20-21 (December, 2010) (.stating, as its first recommendation, that "[t]he Legislature should determine clearly in law who is the owner of medical records.").

[12] *See* Leslie L. Roos *et al.*, *Strengths and Weaknesses of Health Insurance Data Systems for Assessing Outcomes*, *in* INST. OF MED., MODERN METHODS OF CLINICAL INVESTIGATION 47 (Annetine C. Gelijns, ed., 1990), *available at* http://www.nap.edu/openbook.php?record_id=1550 [hereinafter, "IOM, MODERN METHODS"] (discussing the use of administrative data—for example, claims data held by Medicare, Medicaid, and private health insurers—in health research).

[13] Barbara J. Evans, *Congress' New Infrastructural Model of Medical Privacy*, 84 NOTRE DAME L. REV. 585, 596-97 (2009) (discussing the current status of ownership of health data and human tissue specimens).

[14] *See* Susan E. Gindin, *Lost and Found in Cyberspace: Informational Privacy in the Age of the Internet*, 34 SAN DIEGO L. R. 1153, 1195 at n. 231 (1997) (listing several cases where courts have recognized patients' ownership of medical records).

[15] *See* Seth Axelrad, State Statutes Declaring Genetic Information to be Personal Property, *available at:* http://www.aslme.org/dna_04/reports/axelrad4.pdf (listing statutes of Alaska, Colorado, Florida, and Georgia that recognize individual property rights in genetic information).

[16] David L. Silverman, *Data Security Breaches: The State of Notification Laws*, 19 No. 7 INTELL. PROP. & TECH. L. J. 5, 8 (2007).

[17] *See* Paul M. Schwartz, *Property, Privacy, and Personal Data*, 117 HARV. L. REV. 2055, 2076-2094 (2004) (discussing conceptual difficulties in applying a regime of property rights to information in databases). *See also* Silverman, *supra* note 16, at 8 (noting, "Although it is common for businesses contracting with one another to state that one or another of them 'owns' a particular data set, ownership of the contents of a database is a precarious concept in the United States. The contents of a database may be owned in the sense that the database is protectable as a trade secret, but only if the database independently meets the requirements under applicable state law for protection of trade secrets. As with any trade secret, the right may vanish if secrecy is not maintained. Copyright law typically does not provide protection for the contents of a database, because the contents are facts, not expression; if there is protection under copyright law, it is typically limited to the selection and arrangement of the database and does not extend to the content itself.")

[18] M. Bloomrosen & D. Detmer, *Advancing the framework: use of health data—a report of a working conference of the American Medical Informatics Association*, 15 J. AM. MED INFORM. ASSOC. 715-22 (2008) (discussing data stewardship).

[19] U.S. Dep't of Health & Human Servs., Food & Drug Admin., FDA's Sentinel Initiative: Transforming how we monitor the safety of FDA-regulated products, http://www.fda.gov/Safety/FDAsSentinelInitiative/ucm2007250.htm.

[20] JANET M. MARCHIBRODA, eHEALTH INITIATIVE FOUND., DEVELOPING A GOVERNANCE AND OPERATIONS STRUCTURE FOR THE SENTINEL INITIATIVE 34 (2009), *available at* http://www.regulations.gov/#!documentDetail;D=FDA-2009-N-0192-0006.

[21] *Id.* at 21.

entitlements and protections afforded by the HIPAA Privacy Rule and the Common Rule is strikingly similar to what patients would enjoy if they owned their data.[22] Part II of this article explains that both regimes—patient ownership of data, on the one hand, vs. the federal regulatory protections, on the other—provide "pliability-rule"[23] protection that strikes a balance between patient control and the public's need for data access. Both regimes allow unconsented uses of patients' data, and the grounds for unconsented data use are substantively similar under either regime. This similarity suggests that property rights may not be the right locus for reforms. Creating property rights in data would result in a new scheme of entitlements that is substantively similar to what already exists, thus perpetuating the same frustrations all sides have felt with the existing federal regulations.

This article calls for an altogether different debate—a debate about appropriate public use of private data and how best to facilitate it while adequately protecting individuals' interests. However framed, this debate calls for heightened specificity about the nature of the data resources in question, the infrastructures for providing them, and their intended uses.[24] Legal discussions of health information technology sometimes resemble the old parable about blind men discussing an elephant—one palpates the trunk and likens the elephant to a snake; another palpates a leg and likens the elephant to a tree. Apparent disagreements sometimes reflect the fact that we were not actually discussing the same things. Part III of this article explores how different implicit assumptions were at play in the recent debate among Professors Hall,[25] Schulman;[26] and Rodwin[27] over which allocation of initial entitlements to patients' data would best promote access to data for clinical, research, and public health uses.

Part IV challenges a commonly held view that the Health Information Technology for Economic and Clinical Health (HITECH) Act,[28] passed in 2009, did little to promote interoperability and sharing of data for these purposes.[29] The HITECH Act correctly recognized that raw patient data are not the valuable resource; these data acquire value only through the application of infrastructure services. Congress clarified pricing of services that are required to convert raw patient data into valuable resources for research and public health, and Congress authorized data-holders to conduct commercial transactions for sale of those services[30] This approach, which draws on a long tradition of successful American infrastructure development, offers promise in resolving the infrastructure bottlenecks which (rather than the unresolved status of data ownership) have presented the key impediment to data availability.[31]

---

22 *See* discussion *infra* Part II. *See also* Hall & Schulman, *supra* note 8, at 1282 (acknowledging that "the effect of other legal regimes may sometimes resemble property law").

23 *See* Abraham Bell & Gideon Parchomovsky, *Pliability Rules*, 101 MICH. L. REV. 1 (2002) (defining and discussing pliability rules).

24 *See* discussion *infra* Part III.

25 Hall, *supra* note 6.

26 Hall & Schulman, *supra* note 8.

27 Rodwin, *supra* notes 6 and 9.

28 Pub. L. 111-5, Div. A, Title XIII, Div. B, Title IV, 123 Stat. 226, 467 (Feb. 17, 2009).

29 *See, e.g.*, Hall, *supra* note 6, at 635 (pointing out that there is "no legal requirement that funded systems actually interconnect to form a consolidated medical record for each patient"); Rodwin, *supra* note 6, at 595, (discussing the goal of "sharing of patient data for research and public uses" and noting that "HITECH does not appear to authorize creating regulations that can achieve that goal.")

30 *See* discussion *infra* Part IV.

31 *See* discussion *infra* Parts III and IV.

Despite this progress, important problems are unresolved. Since the late 1980s,[32] improved observational methodologies[33] have expanded the array of options for generating scientific and regulatory evidence,[34] making it possible to draw valid scientific conclusions by observing patterns in data that reflect the experiences of large groups of people. Research of this type is set to play a crucial role in the learning healthcare system[35] of the 21st century which will harness electronic records of past healthcare outcomes as a source of evidence to inform future treatment and regulatory decisions[36]—a type of research variously called "records research" or "observational" or "epidemiological" research to distinguish it from interventional (experimental) research such as clinical trials.[37] The slang term "data mining" also is widely used. This article uses the term "health informational" research to distinguish these types of studies from interventional (clinical) research,[38] which was the major workhorse of late-20th-century biomedical discovery.[39]

It has long been understood that informational research and interventional research pose different burdens and risks for participating individuals. Interventional research can cause real physical injury whereas having one's data used can cause undesired disclosure of personal data or dignitary harms such as having one's data used in research of which one does not approve.[40] Less noted, but more important for the present discussion, the individual's decision whether to

---

32 Agency for Healthcare Research & Quality, Outcomes Research Fact Sheet (2000), *available at* http://www.ahrq.gov/clinic/outfact.htm (discussing reasons why observational research flowered after 1980).

33 Brian L. Strom, *Study Designs Available for Pharmacoepidemiology Studies* 17, 21 – 26, in PHARMACOEPIDEMIOLOGY 4th ed. (Brian L. Strom *ed.*, 2005) (discussing the array of scientific methodologies—including observational methodologies that rely on the study of existing data—that are available for studying how people react to drugs).

34 Strom, *id.* at xvi (noting that epidemiological data are now routinely used for regulatory decisions).

35 *See* ROUNDTABLE ON EVIDENCE-BASED MED., INST. OF MED., THE LEARNING HEALTHCARE SYSTEM: WORKSHOP SUMMARY (LeighAnne Olsen *et al*. eds., 2007) [hereinafter, IOM, LEARNING HEALTHCARE], *available at* http://books.nap.edu/openbook.php?record_id=11903 (exploring the potential capabilities of the 21st century "learning healthcare system").

36 Barbara J. Evans, *Seven Pillars of a New Evidentiary Paradigm: The Food, Drug, and Cosmetic Act Enters the Genomic Era*, 85 NOTRE DAME L. REV. 419, 435-39, 485-91 (2010) (describing observational methodologies that rely on the use of existing clinical and administrative data and recent legislation that calls for greater use of these methodologies to ensure drug safety).

37 *See, e.g.*, LAWRENCE M. FRIEDMAN *ET AL*., FUNDAMENTALS OF CLINICAL TRIALS 2 – 5 (3d ed. 1998) (defining clinical trials) ; BENGT D. FURBERG & CURT D. FURBERG, EVALUATING CLINICAL RESEARCH 11- 22 (2d ed. 2007) (discussing clinical trials and their evidentiary strengths and weaknesses) and *id.* at 29 – 37 (describing observational methodologies); Strom, *supra* note 33, at 21-26 (discussing the array of scientific methodologies available for studying drug safety).

38 FRIEDMAN ET AL., *supra* note 37, at 2-5 (discussing interventional research, exemplified by a randomized, controlled clinical trial (RCT) that monitors outcomes prospectively in two groups of people who either were, or were not, subjected to a particular treatment.)

39 Evans, *supra* note 36, at 432 (noting that randomized, controlled clinical trials played a central role in the mid-20th century drug approval paradigm FDA implemented under the 1962 Drug Amendments).

40 *See* Dep't of Health, Educ., & Welfare, Office of the Secretary, Protection of Human Subjects: Institutional Review Boards: Report and Recommendations of the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 43 *Fed. Reg.* 56174, 56181-82 (Nov. 30, 1978) (discussing risks of records research, including risks to confidentiality and the risk that a subject could be put in legal jeopardy if sensitive health information, such as drug abuse records, were disclosed). *See also* David Casarett, Jason Karlawish, Elizabeth Andrews, and Arthur Caplan, *Bioethical Issues in Pharmacoepidemiological Research*, in PHARMACOEPIDEMIOLOGY 4th ed., *supra* note 33, at 587, 588 (noting that "the risks to the subjects of epidemiology research are not the usual health risks of research").

participate has starkly different public impact in the two scenarios. A person's refusal to participate in a clinical trial does not jeopardize the broader clinical research enterprise. Studies still can be completed using other, willing research subjects; only 600 – 3000 people are needed for a typical clinical drug trial.[41] In contrast, nonconsenters in informational research may bias the dataset and reduce its statistical power for everyone.[42] The individual's desire not to participate potentially places other human beings at risk and undermines broader public interests—for example, in public health or medical discovery—in which the individual shares.[43] At the same time, important individual rights are implicated and appropriate respect for these rights somehow must be upheld.

A major challenge in 21st century privacy law and research ethics will be to come to terms with the inherently collective nature of knowledge generation in a world where large-scale informational research is set play a more prominent role.[44] Part V explores two hidden deficiencies of the HIPAA Privacy Rule and Common Rule: First, they conceive the state's police power to use data to promote public health much more narrowly than the police power is conceived in all other legal contexts. This has the effect of thwarting legitimate uses of data to improve the public's health. Second, the provisions for approving nonconsensual research uses of data fail to incorporate any "public use" requirement to ensure that unconsented data uses are justified by a publicly beneficial purpose. As things stand, persons whose health data are used in research have no assurance that the use will serve any socially beneficial purpose at all. These problems, rather than the question of who owns health data, need to be the focus of debate.

II. WHY DATA OWNERSHIP WOULD NOT PROTECT PATIENTS' PRIVACY

A. Nonconsensual Access to Patient Data Under a Property Regime

Data propertization proposals fall into two broad categories: "pro-privacy" proposals that portray private ownership as a way to bolster patients' power to block unwanted uses of their data, and "pro-access" proposals that aim to promote wider availability of data for clinical, research, and public health uses. The pro-privacy proposals rest on a mythical view of private property. Three centuries ago Sir William Blackstone noted how the human imagination is drawn to the idea of property as "that sole and despotic dominion which one man claims and exercises over external things of the world in total exclusion of the right of any other individual in the universe."[45] This idea resonates with the "autonomy *über alles*" strand of privacy advocacy that asserts that the patient's right to control access to health data should trump all other interests,

---

41 COMM. ON THE ASSESSMENT OF THE U.S. DRUG SAFETY SYS., INST. OF MED., THE FUTURE OF DRUG SAFETY 36 (Alina Baciu *et al*. eds., 2007) [hereinafter IOM, FUTURE OF DRUG SAFETY], *available at* http://books.nap.edu/openbook.php?record_id=11750.

42 *See* discussion *infra* Part III.C.

43 *See, e.g.*, Paul Starr, *Health and the Right to Privacy*, 25 AM. J. LAW & MED. 193, 201 (1999). (noting that patients have an interest in privacy of their health records but also have an interest in research and other efforts to improve the medical care available to them).

44 *See, e.g.*, 21 U.S.C.A. § 355(o)(3)(D) (placing the US Food and Drug Administration under a requirement to consider and reject the use of observational studies before the agency can order a postmarket clinical drug trial). *See also* IOM, LEARNING HEALTHCARE, *supra* note 35, at 128 , 130 (discussing the growing use of observational methodologies); Evans, *supra* note 36, at 479-85 (same).

45 WILLIAM BLACKSTONE, 2 BLACKSTONE'S COMMENTARIES ON THE LAWS OF ENGLAND, Chapter the First: Of Property In General 2 *(spelling conformed to modern conventions*), *available at* http://avalon.law.yale.edu/18th_century/blackstone_bk2ch1.asp.

even society's interest in conducting research and public health studies that might save or improve other people's lives. Blackstone was merely describing how people *imagine* property. He himself did not espouse this view,[46] nor has American law ever done so.[47]

Different assets call for different forms of ownership, and proponents of patient data ownership often are indefinite about what they have in mind.[48] Data ownership might, for example, need to look something like the nonexclusive rights riparian owners have in a river that runs by their land—that is, a right to use the river oneself but not to interfere with others' simultaneous uses for fishing and navigation[49] —or like a copyright, which expires after a fixed term of years and which allows "fair use" by others even during that term.[50] Pro-privacy proposals draw on the ideal of property seen in the saying, "One's home is one's castle."[51] Such proposals seek pure "property-rule"[52] protection for patients' data: all uses of data would require the patient's consent on terms defined by the patient, and unconsented uses could be enjoined. [53]

The weakness of pro-privacy proposals is this: having a property right does not ensure property-rule protection. Law recognizes that there are many situations where consensual transactions cannot be relied on as a way of ordering an owner's relations with the larger community.[54] In many circumstances, a property owner only receives "liability-rule" protection[55] and can be forced into nonconsensual transactions in return for a "price" that is externally set, often by a court, legislature, or administrative agency.[56] That price may be zero. The government—when acting under its police power to protect the public's health, safety, morals, or welfare--has broad power to confiscate or interfere with property without paying the owner any compensation. [57] Dating back to colonial times, the state's police power has been used not just to prevent property owners from injuring others, but to pursue broader public welfare

---

46 *Id.* at Chapter the First: Of the Absolute Rights of Individuals 119, available at http://avalon.law.yale.edu/18th_century/blackstone_bk1ch1.asp (recognizing that rights include both rights that are owed to individuals and those that are "due from every citizen, which are usually called civil duties" (*spelling conformed to modern conventions*)).

47 *See* Eric R. Claeys, *Kelo, the Castle, and Natural Property Rights* 35, 40-43, in PRIVATE PROPERTY, COMMUNITY DEVELOPMENT, AND EMINENT DOMAIN (Robin Paul Malloy ed., 2008) (discussing the natural rights theory reflected in 18th and 19th century American takings jurisprudence and how it allowed interference with property rights under certain circumstances).

48 *See, e.g.*, Hall, *supra* note 6, at 663 (calling for an unspecified "right mix and forms of property rights among patients, providers, researchers, and compilers.")

49 Eric R. Claeys, *Takings, Regulations, and Natural Property Rights*, 88 CORNELL L. REV. 1549, 1575 (2003). (2003).

50 Abraham Bell, *Private Takings*, 76 U. CHI. L. REV. 517, 540-42 (2009).

51 Claeys, *supra* note 47, at 35-36 (discussing the popular meaning of the castle metaphor).

52 Guido Calabresi & A. Douglas Melamed, *Property Rules, Liability Rules, and Inalienability: One View of the Cathedral*, 85 HARV. L. REV. 1089, 1092 (1972).

53 *See* Thomas W. Merrill, *The Economics of Public Use*, 72 CORNELL L. REV. 61, 64 (1986) (discussing rights and remedies available under a scheme of property-rule protection).

54 *See* Calabresi & Melamed, *supra* note 52, at 1108-09 (discussing the problems of consensual transactions for compensation of accidents); Bell & Parchomovsky, *supra* note 23, at 8-19 (discussing the evolution of entitlement theory as it bears on the relative merits of consensual and nonconsensual ordering in various circumstances).

55 Calabresi & Melamed, *supra* note 52, at 1092.

56 Bell & Parchomovsky, *supra* note 23, at 3.

57 *See* Merrill, *supra* note 53, at 66 (pointing out, "This is the citizen's plight when the government legitimately exercises its power to tax or its police power…the government [can] take his property without his consent and without compensation.").

objectives for the benefit of the community.[58] "[T]here was no single paradigm of public welfare that confined what we now call the police power. Then, as now, lawmakers pursued a shifting amalgam of goals. … Legislation coercively promoted uses of private land that were viewed as conducive to the community's well-being."[59] If patients owned their data, the data still could be used nonconsensually in public health activities, which long have been viewed as a legitimate exercise of the state's police power.[60]

The state also has eminent domain power to appropriate property for "public use"[61] without the owner's consent, subject to payment of just compensation.[62] The public uses for which property can be taken are quite broad and could include private, commercial research uses of data if data were patient-owned. Takings require "some showing of 'publicness'"of the intended use,[63] and takings that lack the requisite public quality can be enjoined.[64] Public uses traditionally involved placing the taken property under public ownership [65] or transferring it to a private company, such as a utility or railroad, that has obligations to serve the public often but not always for a regulated price. [66] There was never a requirement that the fruits of a taking be made *freely* available to the public; railroads and stadiums built on taken land routinely require users to buy tickets.[67] Though somewhat controversial,[68] modern courts also allow takings that transfer property to new private owners for commercial projects.[69] Such projects need not be

---

58 *See, e.g.*, John F. Hart, *Land Use Law in the Early Republic and the Original Meaning of the Takings Clause*, 94 N.W.U. L. REV. 1099, 1102; (2000). *See also id.* at 1107 (discussing historical uses of the state's police power to require owners to confer positive externalities on the community); William Michael Treanor, *The Original Understanding of the Takings Clause and the Political Process*, 95 COLUM. L. REV. 782 (1995) (same).

59 Hart, *supra* note 58, at 1107.

60 LAWRENCE O. GOSTIN, PUBLIC HEALTH LAW 20-21 (2000); Wendy E. Parmet, *After September 11: Rethinking Public Health Federalism*, 30 J. L. MED. & ETHICS 201 (2002). *See also* Hall, *supra* note 6, at 659 (noting that the government currently requires disclosure of identifiable information for public health purposes under its police power without constitutional objection).

61 *See* Robin Paul Malloy & James Charles Smith, *Private Property, Community Development, and Eminent Domain* 1, 8, in PRIVATE PROPERTY, COMMUNITY DEVELOPMENT, AND EMINENT DOMAIN, *supra* note 47 (discussing the public use requirement).

62 U.S. Constitution, 5th Amendment.

63 Merrill, *supra* note 53, at 61.

64 *Id.*at 68.

65 *See* Bell, *supra* note 50, at 522 (noting, "Public ownership of property is not necessary for just and efficient takings.").

66 *See* Malloy & Smith, *supra* note 61, at 7-8 (discussing evolution of the notions of public use and public purpose over time; *see also* Claeys, *supra* note 47, at 37 (noting that "there are two basic ways to interpret 'public use" as a constitutional term of art": a "use by the public test" that requires the taken property actually be accessible for use by the public and a "public purpose" test that allows the government to redistribute property to promote "general public policies, benefits, or purposes.").

67 Brett M. Frischman, *An Economic Theory of Infrastructure and Commons Management*, 89 *Minn. L. Rev.* 917, 925-26 (2005).

68 *See* Michael Adam Wolf, *Hysteria versus History: Public Use in the Public Eye* 15, 15-20, in PRIVATE PROPERTY, COMMUNITY DEVELOPMENT, AND EMINENT DOMAIN, *supra* note 47 (discussing public reaction to cases that have relied on a broader public purpose test).

69 *See, e.g.*, Poletown Neighborhood Council v. City of Detroit, 410 Mich. 616, 304 N.W.2d 455 (Mich. 1981); Kelo v. City of New London, 545 U.S. 469 (2005).

open to the general public. They may offer only indirect public benefits, such as boosting local tax revenues or aiding urban renewal or land reform.[70]

Eminent domain appears to have been lost on privacy advocates who champion patient ownership of data as a way to halt unconsented, private-sector research use of data. Modern takings doctrine would let privately owned health data be taken for use in academic and commercial research that offers a prospect of developing a beneficial therapy. This is true even if the new therapy, when successfully developed, would be available only to patients who are able to pay for the therapy. It appears doubtful that patients would be entitled to any compensation when their data were taken for research use. Courts construe "just compensation" to mean payment of fair market value—the price the property would fetch in an alternative, consensual sale. There is no compensation for subjective value, such as the emotional attachment an owner has to a particular home, or for undeveloped use rights—what the undeveloped property might have been worth if the current owner had chosen to develop it.[71] There also is no compensation for consequential costs of the taking, such as an owner's moving expenses.[72] These same limitations presumably would apply if patient-owned data were taken for public use in research. When patients wish to let their data "lie fallow" because of privacy concerns, the fair market value of the data arguably is zero. There is no alternative consensual transaction by which to gauge the data's worth. The privacy value of unused data is subjective: it reflects emotional attachment to the data. This is not compensable under modern takings doctrine.[73]

B. Nonconsensual Data Access Under the Existing Federal Regulations

The HIPAA Privacy Rule and the Common Rule offer a framework of patient entitlements and protections that is strikingly similar to what patients would enjoy if they owned their data. Under ordinary circumstances, both regulations require consensual ordering of data access: they require a privacy authorization[74] or informed consent[75] before data can be used. However, both regulations contain exemptions,[76] exceptions,[77] and definitional nuances[78] that shift to a regime of liability-rule protection under certain circumstances.

---

70 *See, e.g.*, Berman v. Parker, 348 U.S. 26 (1954); Hawaii Housing Authority v. Midkiff, 467 U.S. 229 (1984). See also, Bell, *supra* note 50, at 548.

71 *See* Abraham Bell & Gideon Parchomovsky, *The Hidden Function of Takings Compensation*, 96 VA. L. REV. 1673, 1677 (2010) (noting that eminent domain only compensates market value, not emotional attachment or subjective valuation); Claeys, *supra* note 49, at 1600-01, 1632, 1646-47 (noting that takings compensation is based on adverse economic impact in the form of interference with distinct, investment-backed expectations).

72 Malloy & Smith, *supra* note 61, at 8.

73 *See* Hall, *supra* note 6, at 659; Rodwin, *supra* note 6, at 609 (using a different rationale to reach a similar conclusion that no compensation would be owed for takings of de-identified or anonmyzed medical information for public purposes). Hall and Rodwin both argue that patients have no property interest in de-identified/anonymized data; therefore, research uses of such data would not constitute a taking and, hence, no compensation would be owed. My point is different: Even if the data were identifiable or fully identified, and even if the patient had a property interest in the data, and even if research use of the data were deemed to be a taking, it is still true that no compensation would be owed, because takings doctrine does not compensate subjective value and the perceived value of keeping data unutilized is a subjective value.

74 *See* 45 C.F.R. § 164.508 (describing authorization requirements of the HIPAA Privacy Rule).

75 *See* 45 C.F.R. § 46.116 (describing informed consent requirements of the Common Rule).

76 *See* 45 C.F.R. § 46.101(b)-(d) (describing exemptions to the Common Rule).

77 *See* 45 C.F.R. § 164.512 (describing exceptions to the HIPAA Privacy Rule's authorization requirements).

78 *See, e.g.*, 45 C.F.R. § 46.102(d) and (f) (defining the terms "research" and "human subject") and see U.S. Dep't. of Health & Human Servs., Off. For Human Research Protections (OHRP), Guidance on Research Involving Coded Private Information or Biological Specimens 6-7 (Aug. 10, 2004), available at

Certain activities that are considered to have high social value—such as using data for judicial, law enforcement, and public health purposes—are excepted from the usual consent and authorization requirements.[79] Nonconsensual research access is allowed under conditions aimed at reducing privacy risks to the data subjects: if the data have been de-identified,[80] or have been coded in compliance with specified standards,[81] or have been converted to a limited data set.[82] Nonconsensual research uses also are allowed if an Institutional Review Board or privacy board (collectively,"IRB") approves a waiver of the usual consent or authorization requirements.[83] Data supplied to researchers under a HIPAA waiver must meet "minimum necessary"[84] requirements—*i.e.*, no more information can be disclosed than is necessary to accomplish the intended research purpose. However, there is no requirement that the data be de-identified or even coded. In theory, it would be possible to disclose fully identified data under a waiver, if using identified data was necessary to the research and if an IRB found that the other waiver conditions had been met. [85]

While some people object to any nonconsensual use of their data, there is a fairly strong level of public support for police-power uses that protect public health, safety, and welfare.[86] The public also has some degree of comfort with the use of de-identified and other "masked" forms of data[87] despite ongoing concerns about the potential for such data to be re-identified. [88]

---

www.hhs.gov/ohrp/humansubjects/guidance/cdebiol.pdf [hereinafter, "OHRP, 2004 Guidance"] (construing these definitions in a way that causes research with de-identified and coded data and human subjects not to be considered human subject research that requires informed consent under certain circumstances).

79 *See* Barbara J. Evans, *Issue Brief: Appropriate Human-Subject Protections for Research Use of Sentinel System Data*, in FDA SENTINEL INITIATIVE MEETING SERIES: LEGAL ISSUES IN ACTIVE MEDICAL PRODUCT SURVEILLANCE 4 (Engelberg Center for Health Care Reform at the Brookings Institution, March, 2010), http://www.brookings.edu/~/media/Files/events/2010/0308_FDA_legal_issues/Panel%203%20Issue%20Brief.pdf (summarizing the various pathways for nonconsensual use of data under the HIPAA Privacy Rule and Common Rule); Evans, *supra* note 13, at 597, 619-22 (describing in more detail the provisions for nonconsensual data access under the HIPAA Privacy Rule) and at 625-30 (describing nonconsensual access to data under the Common Rule and under the Food & Drug Administration's human-subject protection regulations at 21 C.F.R. pts. 50, 56). *See also* KRISTEN ROSATI, AN ANALYSIS OF LEGAL ISSUES RELATED TO STRUCTURING FDA SENTINEL INITIATIVE ACTIVITIES (2009), *available at* http://www.regulations.gov/#!documentDetail;D=FDA-2009-N-0192-0003.2 (providing a detailed examination of provisions of the Privacy Rule, Common Rule, the Privacy Act, and other relevant laws, such as those governing data on substance abuse, that affect access to data for use in FDA's postmarket drug safety surveillance activities).

80 *See* 45 CFR § 164.514(b) (providing that data can be de-identified, for purposes of HIPAA, by removing 18 specific types of identifiers or by having a statistician certify that the risk of re-identification is "very small"). *See* 45 C.F.R. § 46.102(h) (defining "human subject" in a way that means that research with data is not covered by the Common Rule's consent requirements if investigators do not receive identifying information or interact with the subjects). .

81 *See* OHRP 2004 Guidance, *supra* note 78, at 4 – 6 (discussing permissible coding arrangements under the Common Rule); see 45 C.F.R. § 164.514(c) (allowing coded data to be considered "de-identified" under the HIPAA Privacy Rule if the code-key is subject to certain restrictions on derivation and access).

82 45 C.F.R. § 164.514(e).

83 45 C.F.R. § 164.512(i) [HIPAA waiver provision]; 45 C.F.R. § 46.116(d) [Common Rule waiver provision].

84 45 C.F.R. § 164.514(d).

85 *See* Barbara J. Evans, *Ethical and Privacy Issues in Pharmacogenomic Research*, in Pharmacogenomics: Applications to Patient Care, 2nd Ed. 325, 331 (Howard L. MacLeod et al., eds., 2009).

86 IOM, PRIVACY REPORT, *supra* note 4, at 82.

87 *Id*.

Waivers do not inspire similar levels of public understanding[89] and are subject to ongoing critique, both from research institutions and IRBs that find the waiver provisions cumbersome to apply[90] and from scholars and privacy advocates who view them as an abuse-prone bypass to consent requirements.[91]

The waiver provisions of the HIPAA Privacy Rule and Common Rule are best understood as a regulator-created analogue of privately delegated eminent domain power. These provisions allow private bodies—institutional review boards (IRBs) and privacy boards—to approve nonconsensual research use of data. There is a long history in the United States, dating back to colonial times, of delegating takings power to private parties, such as developers of mill dams and railroads, so that they can take property for specific, socially beneficial uses directly, without having the government act as an intermediary.[92] Private delegations are considered justified when: (1) there are holdouts or other strategic barriers to consensual transactions[93]—that is, when obtaining consent is "impracticable,"[94] which happens to be one of the conditions[95] for granting a waiver; (2) when justice and efficiency are better served by transferring the taken property to a subsequent private owner rather than to the government,[96] as is the case when private-sector research institutions have superior capability to unlock the scientific and public health potential of the data than government agencies possess; and (3) when a repeated pattern of similar transactions would make it administratively burdensome to require governmental involvement.[97] The HIPAA and Common Rule waiver provisions are criticized on various

---

88 *See, e.g.*, FEDERAL TRADE COMM, PROTECTING CONSUMER PRIVACY IN AN ERA OF RAPID CHANGE: A PROPOSED FRAMEWORK FOR BUSINESSES AND POLICY MAKERS 35-38 (Dec., 2010) (warning that the distinction between personally identifiable information and non-identifiable information is increasingly irrelevant in light of the potential for data to be re-identified); Paul Ohm, *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization*, 57 UCLA L. REV. 1701 (2010) (discussing the risks to individual privacy if de-identified data were to be re-identified); Mark A. Rothstein, *Is Deidentification Sufficient to Protect Health Privacy in Research?*, 10 AM. J. BIOETHICS 3 (2010) (same); *But see* Deven McGraw, *Data Identifiability and Privacy*, 10 American Journal of Bioethics 30 (2010) (noting, "Using information in less identifiable form greatly reduces risks to privacy."). *See also* Misha Angrist, *Urge Overkill: Protecting Deidentified Human Subjects at What Price?*, *id*. at 17; Melissa M. Goldstein, *Guiding Deidentification Forward*, *id.* at 27; and Daniel A. Moros and Rosamond Rhodes, *Privacy Overkill*, *id*. at 12 (all pointing out that de-identification of data, though not infallible, nevertheless does afford a considerable degree of privacy protection, such that efforts to improve privacy protections might be better devoted to other, more pressing problems).

89 Evans, *supra* note 13, at 624.

90 *See supra* notes 4, 5.

91 *See* Carl H. Coleman, *Rationalizing Risk Assessment in Human Subject Research*, 46 ARIZ. L. REV. 1, 13–17 (2004) (discussing procedural informality of the Common Rule); Evans, *supra* note 13, at 622-25 (discussing procedural informality of the waiver provisions of the Common Rule and HIPAA Privacy Rule); Evans, *supra* note 85, at 332 (same); Evans, *supra* note 79, at 5 (same).

92 Bell, *supra* note 50, at 517, 549-50, 545. See also, Hart, supra note 58, at 1102.

93 Bell, *supra* note 50, at 534.

94 *See* 45 C.F.R. § 164.512(i)(2)(ii)(B)-(C) (requiring impracticability both of obtaining consent, and of conducting the research without access to the data, before a HIPAA waiver can be granted); 45 C.F.R. § 46.116(d) (requiring impracticability of conducting the research without a Common Rule waiver).

95 45 C.F.R. § 164.512(i)(2) [HIPAA waiver criteria]; 45 C.F.R. § 46.116(d) [Common Rule waiver criteria].

96 Bell, *supra* note 50, at 534.

97 *Id*. at 545, 561-62.

grounds,[98] but the fact remains that there are strong justifications for privately delegating at least some form of power to approve nonconsensual access to data.

Under a property regime, patients' ability to control uses of their data would be quite similar to the substantive entitlements they enjoy under the existing federal regulations. There might, of course, be procedural differences, with the property regime imposing higher "due process costs"[99] on nonconsensual uses of data. Yet high due-process costs are themselves a factor that tends to justify private delegations of takings power. If patients owned their data, it is quite possible that some scheme of private eminent domain—perhaps resembling the waiver provisions of the federal regulations—would need to be fashioned to address the due process costs of securing data access for research activities. It would be consistent with longstanding practice for the government to delegate its eminent domain power to private actors—such as healthcare data environments and research institutions—that are repeatedly involved in transactions to supply data for use in informational research. It is hard to make a case that data ownership would give patients any more control then they now have.

III. WHY PROPOSED OWNERSHIP REGIMES WOULD NOT PROMOTE ACCESS TO DATA FOR RESEARCH AND PUBLIC HEALTH

Turning to the pro-access proposals, the health information privacy community was puzzled recently by a debate in which several of our most admired scholars drew divergent conclusions about the optimal scheme of ownership for patient data. Divergent conclusions are not puzzling in themselves, but they are so when two analyses that embrace similar objectives, similar methodologies, and similar assumptions give rise to the divergence. Both analyses—one by Professors Hall[100] and Schulman;[101] the other by Prof. Rodwin[102]—favor the objective of making health data more widely available for use medical treatment, public health, and research.[103] Both employ resource classification as their methodology—a method in which analysts "classify infrastructure resources as public goods, network goods, natural monopoly, or some combination thereof"[104] to explain "why markets may fail to efficiently supply such goods, and then proceed to analyze the form of institutional intervention by the government to correct the failure."[105] Both state many of the same assumptions: that the use of health data is

---

98 *See supra* note 117 (discussing procedural defects). *See also* discussion *infra* Part V (discussing the absence of a criterion requiring research uses of data to provide public benefits).

99 *See* Merrill, *supra* note 53, at 77 (discussing the procedural complexity of eminent domain, which imposes due process costs in the form of difficulties obtaining legislative authority for a taking, drafting and filing the complaint, serving of process, securing a formal appraisal of the asset's value, and potentially litigating trials and appeals).

100 Hall, *supra* note 6.

101 Hall & Schulman, *supra* note 8.

102 Rodwin, *supra* notes 6 and 9.

103 *See* Rodwin, *supra* note 6, at 586-87 (summarizing the advantages of tapping data from patient records to "improve medical knowledge, patient safety and public health"); Hall, *supra* note 6, at 635-36 (identifying the challenge as being to achieve "an interconnected, automated, networked world where information follows the patient, information-based tools aid in decision making, and population health data can be mined to improve the quality and outcome of care for all").

104 Frischman, *supra* note 67, at 939-40.

105 *Id*. at 929, 939-41.

nonrivalrous;[106] that health data resources generate public goods;[107] that interoperable data systems exhibit network effects;[108] that data-holders such as insurers and healthcare providers enjoy rights somewhat equivalent to ownership of patient data amid the present legal ambiguities;[109] and that control of data resources is highly fractured at the level of these data-holders, leading to a tragedy of the anti-commons.[110]

The authors recommend starkly different policy interventions. Hall and Schulman suggest that it would stimulate market development of interconnected electronic medical records (I-EMRs)[111] if patients had a right to enter commercial transactions to license access to their medical information that is in the custody of insurers, healthcare providers, and other data-holders.[112] In his separate, longer analysis, Hall argues that the U.S. healthcare system's fragmentation is "chronic and deeply entrenched"[113] such that patients' medical information is widely scattered among data-holders who may lack incentives to develop I-EMRs. Patients have rights of access to their own data,[114] but, in Hall's view, lack clear entitlements to transfer these rights on commercial terms to "compilers" that could assemble the patient's scattered data into I-EMRs. "If patients were given ownership of their complete medical treatment and health records, they could license to compilers their rights to that information in a propertized form that could be more fully developed and commercialized"[115]

Rodwin argues "that treating patient data as private property precludes forming comprehensive databases required for many of its most important public health and safety uses" and proposes "that federal law require providers, medical facilities and insurers to report key patient data in anonymized and de-identified form to public authorities, which will create aggregate databases to promote public health, patient safety, and research."[116] He calls for outright public ownership of patients' anonymized data.[117]

Hall's analysis sometimes is characterized as a call for private ownership of data.[118] It should not, however, be confused with the simplistic demands for property rights sometimes

---

[106] Hall, *supra* note 6, at 661 (stating, "Information by its nature is nonrivalrous, meaning it can be used by many people at once without depletion."); *see* Rodwin, *supra* note 6, at 598 (using "public good" to refer to assets for which "an individual's use does not diminish use by another person").

[107] Hall, *supra* note 6, at 643; Rodwin, *supra* note 6, at 598, 618.

[108] *See* Rodwin, *supra* note 6, at 597-98 (stating that patient data exhibit network effects); Hall, *supra* note 6, at 638 (claiming that network effects emerge by connecting medical records).

[109] Hall, *supra* note 6, at 646 (noting that data held by such entities are "out of circulation, if not owned'); Rodwin, *supra* note 6, at 588 (noting that "data holders treat patient data as if it were their private property") and at 593 (asserting "[i]f legislation does not resolve the ownership of data, courts are likely to grant property interests to those who possess [patient] data and preserve the status quo").

[110] Hall, *supra* note 6, at 646-47 (noting that the data holders' "multiple ownership of different pieces of a patient's medical history, making it difficult for anyone to assemble a complete record."); Rodwin, *supra* note 6, at 606 (discussing fracturing of control over patient data at the level of physicians, hospitals, and insurers, and noting a second level of fracturing of control at the level of individual patients).

[111] *See* Hall, *supra* note 6, at 636 (defining I-EMRs).

[112] *Id*. at 638; see also Hall & Schulman, *supra* note 8, at 1283-84.

[113] Hall, *supra* note 6, at 640.

[114] *Id*. at 649-50

[115] Hall, *supra* note 6, at 651.

[116] Rodwin, *supra* note 6, at 589

[117] *See* Rodwin, *supra* note 9, (arguing "The Case for Public Ownership of Patient Data").

[118] *See, e.g.*, Rodwin, *supra* note 6, at 608 ("Professor Mark Hall argues that private ownership can overcome anti-commons problems that block the adoption of integrated EMRs and networks."); AllBusiness, Who should

voiced by privacy advocates with the aim of blocking data access.[119] Hall offers a carefully nuanced analysis, recognizing that precise scope of the patient's entitlement would need to be carefully defined and acknowledging a risk that granting patients additional legal rights could add new strategic barriers in a market that already, in Hall's view, exhibits an anticommons problem at the level of data-holders.[120] Hall and Schulman present their proposal as "one potential solution."[121] Under their proposal, the patient would be able to grant a license to a trusted intermediary, which in turn would be able to: (1) compel the various data-holders to make the patient's medical information available for compilation into an I-EMR (subject, of course, to reimbursing the data-holder's costs of complying with such requests[122]), and (2) act as the patient's agent for purposes of arranging commercial transactions with third parties that desire to use the patient's I-EMR.[123] The patient would control the terms under which the trusted intermediary could license the patient's I-EMR to prospective data users, and patients would have a "nonwaivable right to revoke any permission they give for access or use."[124] This scheme of patient-controlled I-EMRs would differ from familiar ownership of "houses and cars."[125]

Both these analyses are insightful and have advanced the scholarly debate about data ownership, access, and privacy. The comments offered here are intended to build on rather than quibble with these thought-provoking proposals. My principal concern is that neither of the proposals is able to supply data resources of the types needed for public health studies and research. The Hall and Schulman proposal would, however, produce data that could be used to improve patient care. Rodwin's proposal would not provide data useful in patient care. To justify these assertions requires specificity about what the valuable health information resources are, how they are created, and how they are used.

A. Identifying the Valuable Data Resources

Statements such as "Whoever owns patient data will determine whether its benefits can be tapped"[126] overstate the importance of controlling one raw material input to a complex, multistage production process. It is true in the same sense that "Whoever owns iron ore will determine the fate of the shipbuilding industry" is true. Certainly, iron ore is a critical input to building a ship, but it is just one of many factors that influence development of the essential infrastructures for turning iron ore into the valuable asset, steel, and for turning the steel into ships. Optimizing ownership arrangements for iron ore may or may not lead to a thriving shipbuilding industry.

In ordinary usage, terms like "medical information" and "health data" can refer to several different information resources:

---

own electronic medical records?, http://www.allbusiness.com/print/13276486-1-22eeq.html (labeling the Hall/Schulman analysis as "the case for private ownership" and presenting it as "the opposing view" to Rodwin's "case for public ownership of data").

[119] *See* discussion *supra* at Part II.

[120] Hall, *supra* note 6, at 646-47.

[121] Hall & Schulman, *supra* note 8, at 1284.

[122] Hall, *supra* note 6, at 650 (noting that the patient's access to patient records held by HIPAA-covered entities is subject to payment of fees to cover the costs of preparing and copying the records, and offering that a "possible solution for the fee problem is insurance reimbursement.")

[123] *Id*. at 660-61.

[124] *Id*. at 661.

[125] Hall & Schulman, *supra* note 8, at 1282.

[126] Rodwin, *supra* note 6, at 587.

1. ***Records of a patient's various encounters with the healthcare system ("encounter-level patient data").*** These are the records of a patient's various encounters with the healthcare system, typically in the form of paper charts and records or electronic files stored by numerous healthcare providers, payers, clinical laboratories, pharmacies and other sellers of medical products with which the patient has done business in the course of receiving health care. Hall's "medical information"[127] and Rodwin's "patient data"[128] often refer to encounter-level patient data. Hall's electronic health records (EMR) are electronic records of a patient's encounters with a single healthcare site, such as a hospital or a physician's office.[129]
2. ***The patient's longitudinal health record (LHR).*** An LHR compiles a patient's encounter-level data from disparate sources to form an extended chronological record that tracks the patient's illnesses, treatments, and outcomes over multiple encounters with the healthcare system. Hall and Schulman refer to these as "longitudinal patient records"[130] or "a consolidated medical record for each patient."[131] Hall's I-EMRs would produce LHRs by "facilitat[ing] the compilation of a patient's entire medical treatment and health history from among multiple independent records holders."[132]
3. ***Longitudinal, population health data (LPHD).*** LPHD gathers LHRs from many patients to create a dataset that reflects the healthcare experiences of a large number of people.[133] Hall's trusted intermediaries would be able to enter transactions that gather patient-specific I-EHRs together to form LPHD. Rodwin's "national patient database" is intended to generate LPHD.
4. ***Unbiased LPHD.*** This is a subcategory of LPHD that, in addition to the characteristics just described, has a special attribute: the LPHD provides a representative sample[134] of a larger population about which researchers or public health authorities (together, "investigators") are trying to draw conclusions. Individual LHRs that have been included in the LPHD constitute a representative sample of a larger population of interest. Studying unbiased LPHD will let investigators draw scientifically valid conclusions that are generalizable to the larger population.[135] LPHD obviously is unbiased if it includes

---

127 Hall, *supra* note 6, at 646 (noting, "Although medical information is not property, medical records are.")

128 *See, e.g.*, Rodwin, *supra* note 6, at 589 (calling for providers, medical facilities, and insurers to be required to report patient data in anonymized form to public authorities) and at 595 (stating the belief that "tapping the real potential for patient data for secondary uses requires that it be aggregated into a national database").

129 *See, e.g.*, Hall, *supra* note 6, at 643-44 (referring to a hospital's EMR).

130 Hall & Schulman, *supra* note 8, at 1284.

131 Hall, *supra* note 6, at 635.

132 *Id.*at 651.

133 Evans, *supra* note 13, at 592.

134 *See* David M. Eddy, Should We Change the Rules for Evaluating Medical Technologies?, in IOM, MODERN METHODS 117, 124-25, *supra* note 12 (discussing various types of bias that can occur in scientific studies and their impact on generalizability of results).

135 *See infra* note 153 and accompanying text for discussion of empirical studies demonstrating biases that can result when inclusion of individuals' data into a dataset is predicated on informed consent. *See* FRIEDMAN ET AL., *supra* note 37 *and* FURBERG, *supra* note 38 (for general discussions of problems that can affect data quality, including biases that can undermine the generalizability of results). *See* Brian L. Strom, *Sample Size Considerations for Pharmacoepidemiology Studies* 29, 29 - 36, in PHARMACOEPIDEMIOLOGY, *supra* note 33 (discussing the sample sizes required for various types of health informational research) and Suzanne L. West, Brian L. Strom, and Charles Poole, *Validity of Pharmacoepidemiologic Drug and Diagnosis Data*, *id*. at 709, 709-766 (discussing problems with data quality in health informational research) .

LHRs for *all* members of the larger population in question, for example, if it includes data for all Americans or data for everybody in the world. Rodwin's concept of "national, longitudinal health data"[136] is a form of unbiased LPHD. Unbiased LPHD also can be created using smaller samples of individuals' LHRs, so long as a random, representative sample is obtained. Hall's statement that "population health data can be mined to improve the quality and outcome of care for all"[137] presumes the use of unbiased LPHD.

Encounter-level patient data and individual LHRs are useful for purposes of treating the individual patient. For example, LHRs can answer questions about a patient's medical history that are relevant to the current treatment encounter, or they can notify physicians about treatments that other care providers have administered to the same patient for the same illness so that duplicative or conflicting treatments can be avoided. Encounter-level patient data and individual LHRs, which include data for just one person, have limited or no direct use as resources for public health studies and research; however, they are raw material from which useful data resources for research and public health can be derived. The useful resource for public health and research activities is LPHD.[138] For the vast majority of research and public health studies, there is a further requirement to use unbiased LPHD so that valid, generalizable scientific conclusions can be drawn. There are some research and public health applications that can work with biased LPHD (which, for example, may be useful in preliminary studies to generate hypotheses for later, more rigorous study). In general, however, the demand for LPHD for use in research and public health is demand for unbiased LPHD.

B. The Problem of Linking Data Across Healthcare Data Environments

Rodwin's analysis captures an essential truth about the need for nonconsensual ordering of access to data for certain research and public health purposes.[139] However, his proposed policy solution (requiring data-holders to report anonymized encounter-level patient data to a publicly owned national database) runs up against a serious technical problem: It is impossible—not merely costly or difficult, but impossible—to make longitudinal health records out of encounter-level patient data that have been anonymized.[140] Linking data longitudinally to create LHRs requires at least some identifying information to establish whether the encounter-level data

---

[136] Rodwin, *supra* note 6, at 587.

[137] Hall, *supra* note 6, at 635-36.

[138] *See* U.S. Food & Drug Admin., U.S. Dep't of Health and Human Servs., Proceedings, Sentinel Network Public Meeting 74 (Mar. 8, 2007) [hereinafter FDA, March 8 Proceedings], http://www.fda.gov/ohrms/dockets/dockets/07n0016/07n-0016-tr00002.pdf (statement of Dr. Clement McDonald) (discussing the importance of longitudinal population health data in research and noting the difficulties of linking data from disparate data sources). *See also* U.S. Food & Drug Admin., U.S. Dep't of Health and Human Servs., Sentinel Network Public Meeting 51-56 (Mar. 7, 2007) [hereinafter FDA, March 7 Proceedings] http://www.fda.gov/ohrms/dockets/dockets/07n0016/07n-0016-tr00001.pdf (statement of Dr. Marc Overhage) (discussing the importance and difficulty of linking data longitudinally. *See also* PHARMACOEPIDEMIOLOGY, *supra* note 34, at pt.III (containing a series of articles describing the types of data that are useful in various types of pharmacoepidemiological studies.

[139] *See* discussion *infra* this section.

[140] *See* FDA, March 7 Proceedings, *supra* note 138, at 51-56 (statement of Dr. Overhage) (discussing the process of linking data longitudinally and noting the necessity for some sharing of identifiable information) and FDA, March 8 Proceedings, *supra* note 138, at 74 (statement of Dr. MacDonald) (discussing the need for sharing of identifiable information to accomplish linkage). *See also* Evans, *supra* note 13, at 594-96, 606.

received from various data-holders relate to the same patient.[141] If the goal is to make an anonymized LHR, the order of operations matters: first, identifiable encounter-level patient data are linked together to make an identifiable LHR; then the identifiable LHR is anonymized. The linkage must precede the anonymization.

Rodwin calls for encounter-level patient data to be "anonymized or de-identified"[142] at the source by each data-holder, which then would report the anonymized data to a centralized, national database that would somehow "create aggregate databases to promote public health, patient safety, and research."[143] His proposal would let each data-holding facility report its data in coded form—that is, with an individual tracking number that allows data from the patient's subsequent encounters with that facility to be correlated with data previously reported.[144] Unfortunately, these facility-level tracking numbers (coding[145]) would not allow a patient's data be linked with data from other facilities where the patient has received care.[146] Each facility has its own coding system (tracking number). Facilities would need to share their code keys in order to establish that Tracking Number 13275 at Dr. Brown's office and Tracking Number 999345 at Central Hospital both refer to Mary Smith. Unfortunately, sharing of code keys amounts to sharing of identifiable data.[147] This is prohibited under Rodwin's proposal, which only allows reporting of de-identified patient data.

Suppose Mary Smith visits Dr. Brown for arthritis pain and is prescribed rofecoxib (Vioxx). Two months later, she is admitted at the emergency department of Central Hospital for a stroke. Six months later, she visits Dr. Brown again for a skinned knee. Under Rodwin's proposal, the national database would contain the following information: Dr. Brown treated anonymous patient 13275 for arthritis, prescribed an oral steroid; this same patient (identified by Dr. Brown's tracking number) was later treated for a skinned knee. Central Hospital treated anonymous patient 999345 for a stroke. There is no way to link Mary's data from Dr. Brown and Central Hospital into a complete LHR unless they divulge that tracking numbers 13275 and 999345 both refer to Mary Smith. If the data in the national database are "mined" for information about Vioxx safety, investigators will see a possible association between taking Vioxx and skinning one's knee, but they will not be able to detect the possible association between taking Vioxx and having a stroke.

Rodwin did not discuss how the national database would "aggregate"[148] the encounter-level patient data it receives. This much can be inferred: The proposed national database would not be able to compile LHRs for each patient, since it would lack the identifiable information necessary for linking the patient's encounters across multiple facilities and data-holders. Unable to create LHRs, the national database could not produce the "[n]ational, longitudinal patient

---

141 *Id.*

142 Rodwin, *supra* note 6, at 589.

143 *Id*.

144 *Id*. at 615.

145 *See* Evans, *supra* note 13, at 619-31 (discussing coding of data and its significance under the HIPAA Privacy Rule, the Common Rule, and the FDA human-subject protection regulations at 21 C.F.R. pts. 50, 56).

146 *See* Barbara J. 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. 67, 76-77 (2010) (providing diagrams and a discussion of the problems of linking data across multiple healthcare data environments).

147 45 C.F.R. § 164.514(c)(2) [HIPAA Privacy Rule]; OHRP, 2004 Guidance, *supra* note 78, at 4-5 [Common Rule].

148 Rodwin, *supra* note 6, at 589.

data"[149] (LPHD) that are needed for public health activities and research. The proposed national database would merely contain a second, anonymized copy of the same fragmented, unlinked, disorganized data that already exist. Unless encounter-level patient data are reported to the government in identifiable form—a policy that is far more problematic than the one Rodwin has proposed—it is difficult to see how a national database would add any value.

C. The Problem of Consent Bias

The proposal by Hall and Schulman solves the data-linkage problem by relying on consensual ordering.[150] The patient could authorize a trusted intermediary to obtain identifiable encounter-level data, which then could be linked to create the patient's LHR. The resulting LHR would be useful for purposes of the patient's own care.[151] It is not at all clear, however, that this proposal could generate data to support research and public health activities. The problem relates to the consensual ordering inherent in Hall and Schulman's scheme of patient-controlled health records.

The Hall/Schulman proposal allows the trusted intermediary to use a patient's LHRs to form LPHD and to license the LPHD to third-party users—*but only on terms controlled by the patient*.[152] Multiple empirical studies[153] have documented that people who are willing to consent to letting their data be used in research are different *medically* from the population at large. The underlying reasons are not well understood, but the impact is clear: conditioning the creation of LPHD on patient consent produces datasets that are unreflective of the general population, thus biasing study results.[154] Similar problems also exist outside the biomedical context. Burstein has explored how efforts to reduce the vulnerability of our nation's critical information infrastructures are impeded by researchers' lack of access to realistic data about people's Internet usage patterns and electronic communications, including content and non-content information.[155] The Omnibus Crime Control and Safe Streets Act of 1968,[156] as amended by Electronic Communications Privacy Act (ECPA) of 1986,[157] would let entities like Internet service providers that possess this information share it with researchers if the affected Internet users

---

149 *Id*. at 587.

150 *See* Hall & Schulman, *supra* note 8, at 1284 (calling for "patient controlled health records.")

151 *See* discussion *infra* this section.

152 Hall, *supra* note 6, at 660-61.

153 For discussion of biases caused by the non-random distribution of persons willing to provide consent, *see generally* B. Buckley *et al*., *Selection Bias Resulting from the Requirement for Prior Informed Consent in Observational Research: A Community Cohort of People with Ischaemic Heart Disease*, 93 HEART 116 – 20 (2007); Casarett *et al*., *supra* note 40, at 587, 593-94; Khaled El Emam *et al*., *A Globally Optimal k-Anonymity Method for the De-identification of Health Data*, 16 JOURNAL OF THE AMERICAN MEDICAL INFORMATICS ASSOCIATION 670 – 82 (2009); S.J. Jacobsen *et al*., *Potential Effect of Authorization Bias on Medical Record Research*, 74 MAYO CLIN. PROC. 330 – 38 (1999); J. V. Tu *et al*., *Impracticability of Informed Consent in the Registry of the Canadian Stroke Network*, 350 NEW ENGL. J. MED. 1414 – 21 (2004); S. H. Woolf *et al*., *Selection Bias from Requiring Patients to Give Consent to Examine Data for Health Services Research*, 9 ARCH. FAMILY MED. 1111 – 18 (2000). *See also* IOM, PRIVACY REPORT, *supra* note 4, at 209-14 (surveying studies of consent and selection bias).

154 *Id.*

155 Aaron J. Burstein, *Amending the ECPA to Enable a Culture of Cybersecurity Research*, 22 HARV. J.L. & TECH 167, 170-71, 184 - 94 (2008).

156 Pub. L. No. 90-351, §§ 801–804, 82 Stat. 197, 211–25 (codified as amended at 18 U.S.C. §§ 2510–2522).

157 Pub. L. No. 99-508, 100 Stat. 184 (codified as amended in scattered sections of 18 U.S.C.).

consent.[158] The prospects seem dim that large numbers of people would consent to the release of their private information—including, possibly, the content of their e-mails—to researchers. Even if a few altruistic and unembarrassed souls are willing to let researchers use their e-mails, such people may not provide a representative sample of Internet users more generally. Consenters may not include the cyberterrorists that researchers were hoping to study.

Because of consent bias, Hall and Schulman's patient-controlled I-EMRs can generate patients' LHRs but cannot produce the high-quality, unbiased LPHD that researchers and public health officials need in order to draw scientifically valid, unbiased conclusions. The demand for biased LPHD is questionable. The Hall/Schulman proposal envisions that licensing fees paid by third-party data users would help finance the informational infrastructure for compiling patients' LHRs.[159] Given the low quality of the LPHD a patient-controlled system can generate, demand from research and public health users may be very limited; their licensing fees may not be a reliable source of funding for the system.[160]

Where Rodwin's analysis excels is in its exposition of supply-side factors that call for nonconsensual ordering of access to data for research and public health applications.[161] Hall and Rodwin both acknowledge that diffusion of control among multiple data-holders can give rise to a tragedy of the anticommons.[162] Rodwin explores an additional tragedy of the anticommons that arises when control over data is diffused at the level of individual patients.[163] Likening the problem of assembling "comprehensive patient databases" to the problem of assembling contiguous parcels of land for real-estate development, he explores strategic barriers that make consensual access to data unworkable.[164]

Rodwin's arguments aptly capture the impacts of choosing consensual vs. nonconsensual ordering of data access. He frames this discussion as a comparison of private and public ownership, and this framing is unfortunate at times. For example, the statement "treating patient data as private property precludes forming comprehensive databases required for many of its most important public health and safety uses,"[165] is true only if property rights are modeled as conferring property-rule protection (pure consensual ordering). It discounts the possibility that the needed public access to privately owned data could be obtained nonconsensually through exercises of the police or eminent domain powers.[166] The need for nonconsensual access to data does not necessarily imply a need for public ownership. As Bell has remarked when discussing takings that transfer property into public ownership: such actions are "warranted only where two issues are resolved in favor of the government: (1) the government is the preferred owner for

---

158 Burstein, *supra* note 155, at 186.

159 *See* Hall, *supra* note 6, at 646 (envisioning that "propertizing medical information could stimulate increased flow of medical information into more useful forms by giving stakeholders rights that they can license or sell.").

160 Note that Hall and Schulman never expressly claimed that their proposed scheme would produce data suited to research and public health uses; they may have envisioned I-EMRs primarily as a tool to improve clinical care. *See* Hall, *supra* note 6, at 650 (suggesting that health insurers might be a source to help cover the costs of generating patient's interconnected EMRs—a notion that seems to presume I-EMRs would be used in clinical care rather than in research and public health uses).

161 *See* Rodwin, *supra* note 6, at 603-06.

162 *Id*. at 606; Hall, *supra* note 6, at 647-48.

163 Rodwin, *supra* note 6, at 606.

164 *Id*. at 607. *See also* Hall, *supra* note 1, at 647 (invoking the land-assembly analogy to describe strategic barriers in getting multiple data holders to cooperate to assemble a patient's complete longitudinal health record).

165 Rodwin, *supra* note 6, at 58 .

166 *See* discussion *supra* Part II.

reasons of justice or efficiency, and (2) coercion is the preferred transfer mechanism."[167] The same two questions loom here: is consensual access to data warranted and, if so, is the government the best data-owner?

Rodwin has made the case for some form of nonconsensual access to data for at least some types of research and public health activities. Public ownership would require further justification. Federal agencies such as the Department of Health and Human Services (HHS), which would own the data under Rodwin's proposal,[168] face an additional layer of privacy regulation under the Privacy Act,[169] on top of the HIPAA Privacy Rule and Common Rule which also would apply. This and other potential disadvantages of public ownership need to be carefully weighed, even if one is prepared to accept that HHS has the resources to construct and operate a mega-database containing duplicate copies of all of the health data in the United States.

D. The Role Infrastructure and Demand-Side Factors

Hall's proposal does not embrace any particular system architecture for implementing I-EMRs. He merely states that "[t]he primary barriers are not technological"[170] and turns to analysis of the perceived legal barriers. Rodwin's analysis implicitly assumes that a centralized database is necessary in order to assemble encounter-level patient data into LHR and LPHD: He asserts that "tapping the real potential for patient data for secondary uses requires that it be aggregated into a national database."[171] His preference for public ownership may have been influenced by the assumption that public access to, and use of, data requires actual possession of the data in a centralized database.

There are multiple system architectures that can convert encounter-level patient data into valuable data resources for research and public health.[172] Government-mediated transfers[173] are not the only mechanism for moving data resources into the hands of clinicians and investigators. It is true that, in the past, informational research typically was performed by gathering data into one large, central database where the data analysis was performed.[174] The modern trend is to use distributed data networks instead.[175] Centralized databases worked satisfactorily in the days—not so long ago—when a "large-scale" observational study might have involved mere tens to hundreds of thousands of records. Today, however, large-scale studies may use records for tens to hundreds of millions of persons.[176] For example, the Food and Drug Administration Amendments Act of 2007[177] (FDAAA) calls for pharmacoepidemiological[178] studies of

---

167 Bell, *supra* note 50, at 534.

168 *See* ROSATI, *supra* note 79, at 5.

169 5 U.S.C. 552a.

170 Hall, *supra* note 6, at 636.

171 Rodwin, *supra* note 6, at 595.

172 Carol C. Diamond, Farzad Mostashari, and Clay Shirky, *Collecting and Sharing Data For Population Health: A New Paradigm*, 28 HEALTH AFFAIRS 454, 456 (2009).

173 *See* Rodwin, *supra* note 6, at 589 ("Public authorities should also make this data available for private entities to develop data-derived services, subject to public oversight.")

174 Diamond *et al.*, *supra* note 172, at 456.

175 Richard Platt *et al, The New Sentinel Network—Improving the Evidence of Medical-Product Safety*, 361 NEJM 645-47 (2009). *See also*, Diamond *et al, supra* note 172, *at* 460.

176 *See* Evans, *supra* note 146, at 73-74 (describing several multimillion-person pharmacoepidemiological data networks now under development).

177 The Food and Drug Administration Amendments Act of 2007 (FDAAA), Public Law 110-85, 121 Stat. 823 (September 27, 2007), codified at scattered sections of 21 U.S.C.

postmarket drug safety that will employ health data for 100 million persons.[179] FDA is meeting this mandate by developing the Sentinel system,[180] and its pilot Mini-Sentinel[181] system already incorporates data for 60 million persons.[182] Multimillion-person pharmacoepidemiological networks also are being developed in Canada,[183] the European Union,[184] and Japan.[185] *These systems have not required any clarification of data ownership and they did not require creation of centralized databases*. They all rely on distributed network architectures.

Under a distributed network approach, people's health data remain in their current locations (for example, in insurers' administrative databases or in providers' clinical databases) and are not physically transferred to a central location for storage and analysis.[186] The participating data-holders are linked together virtually.[187] Under one design, parties wishing to access and use data send queries to the data-holders.[188] Suppose, for example, that an investigator wishes to study whether taking statins may be associated with the muscle-wasting condition known as rhabdomyolysis. The investigator would send queries to the various data-holders (for example, "Please locate records for any person in your data system who: (1) has ever

---

178 *See* Brian L. Strom, *supra* note 33, at 3 (defining pharmacoepidemiology as "the study of the use of and the effects of drugs in large numbers of people").

179 FDAAA § 905(a); 21 U.S.C. § 355(k)(3)(B)(ii) (setting targets of 25 million persons by July 2010 and 100 million by July 2012); *see also* 21 U.S.C. § 355(k)(3)(C) (describing the new "postmarket risk identification and analysis system").

180 U.S. DEPT. OF HEALTH & HUMAN SERVS., FOOD & DRUG ADMIN., THE SENTINEL INITIATIVE 1 (2008), http://www.fda.gov/oc/initiatives/advance/reports/report0508.pdf (discussing the goals and structure of the Sentinel data network). *See also* U.S. Dept. of Health & Human Servs., Food & Drug Admin., FDA's Sentinel Initiative, http://www.fda.gov/oc/initiatives/advance/sentinel/ (providing information about the current status of Sentinel System development.

181 U.S. Dept. of Health & Human Servs., Food & Drug Admin., FDA Awards Contract to Harvard Pilgrim to Develop Pilot for Safety Monitoring System (Jan. 8, 2010), http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm196968.htm.

182 Rachel E. Behrman, Joshua S. Benner, Jeffrey S. Brown, Mark McClellan, Janet Woodcock, and Richard Platt, *Developing the Sentinel System—A National Resource for Evidence Development*, 36 NEW. ENG. J. MED. 498 (2011).

183 *See* Canada Institutes of Health Research (CIHR), In Brief: The *Drug* Safety and Effectiveness Network (DSEN), http://www.cihr-irsc.gc.ca/e/39389.html. *See also* Health Canada, Medicines that Work for Canadians: Business Plan for a Drug Effectiveness and Safety Network (2007), http://www.hc-sc.gc.ca/hcs-sss/pubs/pharma/2007-med-work_eff/index-eng.php.

184 *See* European Network of Centres for Pharmacoepidemiology and Pharmacogvigilance (ENCePP), http://encepp.edu; European Risk Management Strategy, Two-Year Work Programme (2008-09), http://www.emea.europa.eu/pdfs/human/phv/28008907en.pdf (describing the ENCePP data network). *See also* Welcome to the EU-ADR Website, http://www.alert-project.org/ (describing the EU-ADR data network). *See also* EMEA-coordinated PROTECT project has been accepted for funding by the Innovative Medicines Initiative Joint Undertaking, Pharmanews (April 30, 2009), http://www.pharmanews.eu/emea/197-emea-coordinated-protect-project-has-been-accepted-for-funding-by-the-innovative-medicines-initiative-joint-undertaking (describing the PROTECT network).

185 Kaoru Misawa, Director, Office of Safety, Pharmaceuticals and Medical Devices Agency (PMDA), *Sentinel Initiative in Japan: Utilization of Electronic Health Information in Pharmacovigilance*, 9th Kitasato University-Harvard School of Public Health Symposium (11-12 September, 2009).

186 Deven McGraw, Kristen Rosati & Barbara Evans, *A Model for Advancing Public Health and Protecting Privacy*, PHARMACOEPIDEMIOLOGY & DRUG SAFETY (*forthcoming*, 2011).

187 *See* Evans, *supra* note 146, at 75-78 (discussing distributed architectures).

188 *Id.* at 77, fig. 2 (showing a distributed network query structure that provides for longitudinal linkage of data across participating data environments via a trusted intermediary).

taken statins, or (2) has ever suffered from rhabdomyolysis."). Records for such patients could be conveyed, in identifiable form, to a network coordinating center (trusted intermediary) that would perform longitudinal linkage of the data received from the various data-holders. This linkage would make it possible to identify patients who *both* took statins *and* suffered rhabdomyolysis, even if the records of these two occurrences are scattered among multiple data-holders. The trusted intermediary would use the linked data to compile lists of patients who took statins with and without subsequently developing rhabodmyolysis. These lists then could be de-identified and conveyed to the investigator for use in the study.

Distributed architecture offers a number of advantages over central databases.[189] Obviously, it avoids the need to invest in duplicative storage capacity, since data reside with their original data-holders and are not redundantly stored at a central location. It offers advantages in privacy and data security, since the data continue to reside behind the privacy firewalls of their original data-holders, with movements of data minimized to what is necessary to respond to specific queries (as opposed to moving all data to a central repository in anticipation of unspecified future uses).[190] Perhaps the most important advantage, in terms of data quality, is that distributed networks let the encounter-level data be interpreted and processed by the data-holder's own personnel, who regularly work with the data and are familiar with its quirks.[191] Data-holders do not all use standardized record formats.[192] Different healthcare providers and insurers describe the same medical condition in different ways, just as law professors use different terminology to refer to similar concepts (for example, LHR, I-EMR, complete patient record, longitudinal patient data). To answer a simple question, such as whether a patient actually had rhabdomyolysis, requires familiarity with how relevant data have been recorded in the particular data system. The President's Council of Advisors on Science and Technology (P-CAST) is pessimistic that a standard record format will ever emerge: "[W]e believe that any attempt to create a national health IT ecosystem based on standardized record formats is doomed to failure. … With so many vested interests behind each historical system of recording health data, achieving a natural consolidation around one record format for any particular subset of data would be difficult, if not impossible."[193] The notion that a national database operator could make sense of raw, encounter-level patient data reported in disparate formats is fanciful.

When data are stored in multiple formats, assembling LHRs and LPHD requires two types of inputs: (1) encounter-level patient data, and (2) services.[194] In a distributed data network, the data-holders supply both.[195] To respond to a data request, personnel of the data-holder must locate and interpret which data in their systems are relevant to the particular query, retrieve the data, and convert the information to a common format that will allow data from

---

189 *See* Diamond *et al.*, *supra* note 172; Platt *et al.*, *supra* note 205 (discussing these advantages).

190 *Id*.

191 *Id*.

192 EXECUTIVE OFFICE OF THE PRESIDENT, PRESIDENT'S COUNCIL OF ADVISORS ON SCIENCE AND TECHNOLOGY, REPORT TO THE PRESIDENT: REALIZING THE FULL POTENTIAL OF HEALTH INFORMATION TECHNOLOGY TO IMPROVE HEALTHCARE FOR AMERICANS: THE PATH FORWARD 39 (December, 2010) [hereinafter, "P-CAST REPORT"]

193 *Id*.

194 *See* Evans, *supra* note 146, at 86-90 (discussing the types of infrastructure that FDAAA envisions will be necessary to support operations of FDA's Sentinel System).

195 *Id*.

multiple data-holders to be combined. Encounter-level patient data is transformed into valuable information resources (LHR and LPHD) through the addition of services.

This fact has important implications. It no longer can be said that "[w]hoever owns patient data will determine whether its benefits can be tapped"[196] Tapping the benefits requires both data and services, and control over data is unavailing without the services. It also is not true that health data resources are nonrivalrous.[197] It is probably fair to say that encounter-level patient data are nonrivalrous, since multiple users could use these data without using them up—assuming these data were useful, which generally they are not except in patient care. LHR and LPHD, which are quite useful, are subject to potential supply constraints: there is a finite supply of the services needed to make them. Data-holders do not have unlimited personnel and data processing resources to respond to queries. Preparing LPHD to respond to one query may diminish the availability of LPHD for another query. The valuable information resources for clinical, research, and public health applications are LHR and LPHD, and these can only be supplied by a constrained infrastructure. These resources are only partially nonrivalrous—that is, they are nonrivalrous only within capacity constraints.[198]

The fact that the necessary services are in finite supply (and are costly) has ramifications for system design. A key design decision is whether the system needs to be able to produce LHR and LPHD ahead of demand, as opposed to satisfying demand after it arises. The answer depends on whether the planned applications—clinical care, research, and public health studies—are latency-sensitive.[199] The concept of latency (delay) has been a concern in discussions of Internet policy.[200] Some Internet applications are latency-sensitive—that is, small delays in delivery of information will disrupt their functionality—while others are latency-insensitive. "Consider that it doesn't matter much whether an email arrives now or a few milliseconds later. But it certainly matters for applications that want to carry voice or video."[201]

Discussions of health information policy often lump all uses together and assume that the optimal infrastructure for supplying data resources for one use would be optimal for others as well. In fact, there are important distinctions. Clinical uses of LHRs are potentially latency-sensitive: clinicians treating a patient in the emergency department cannot afford to wait for compilation of the patient's LHR. On the other hand, the use of LHRs in scheduled clinical care may not be latency-sensitive: when a patient makes a doctor's appointment, a request could be made to compile the patient's LHR for delivery on the date of the scheduled appointment. Many research and public health uses of LPHD are latency-insensitive: it does not destroy the validity of a study if it takes a few days or weeks to supply the necessary data resources.

For latency-sensitive applications, data resources need to be compiled ahead of the demand for them. Patient-controlled I-EMRs, such as those proposed by Hall and Schulman, are thus a potentially useful tool for clinical care. Patients can request compilation of their LHRs in advance so that they will be available in emergencies and then periodically update their LHRs on an ongoing basis. For latency-insensitive applications, such as most research and public health studies, compilation can be deferred until there is identified demand. This distinction affects the

---

196 Rodwin, *supra* note 6, at 587.

197 *See* Hall, *supra* note 6, at 661 stating, ("Information by its nature is nonrivalrous").

198 *See* Frischman, *supra* note 67, at 942 (defining partially nonrivalrous resources).

199 *See* Tim Wu, *Network Neutrality, Broadband Discrimination*, 2 J. ON TELECOMM. & HIGH TECH. L. 141, 148 (2003) (defining and discussing the impact of latency on Internet applications).

200 *Id*., *see also* Frischman, *supra* note 67, at 1008-10.

201 Wu, *supra* note 199, at 148.

required system design and can drastically affect system costs when, as here, compiling the information resources requires inputs of scarce (and costly) services. There would be little advantage—and an enormous cost disadvantage—in developing a centralized, national database containing every person's pre-compiled LHR. Compiling information resources (LHRs and LPHD) in anticipation of all conceivable research and public health uses may be as ill-advised as it would be to manufacture a set of false teeth for every American in anticipation that they may eventually need them; it makes more sense to wait to see who ultimately needs them and then manufacture the teeth.

A distributed architecture can respond to queries as they occur. This offers important economic advantages in latency-insensitive research and public health applications. In the future, it is hoped that development of new infrastructure will reduce the latency itself—in other words, reduce the delays associated with locating relevant patient data and converting them to a consistent format for assembly into LHRs and other useful data resources. At present, these services are labor-intensive. The recent P-CAST report calls for creation of a universal exchange language and infrastructure to facilitate assembly and sharing of patient data across data-holders.[202] Data-holders would continue to operate a variety of systems, including the old legacy systems in operation today and new recordkeeping systems and formats.[203] The "syntax for the universal exchange language will be some kind of extensible markup language (an XML variant, for example) capable of exchanging data from an unspecified number of (not necessarily harmonized) semantic realms." Individual data elements—such as a person's X-ray a clinical observation about the patient—would be annotated with metadata tags that contain enough identifying information to let the patient's records be located and that record information about the patient's privacy preferences and about the provenance of the data (such as what healthcare providers were involved and what type of test or equipment they used).[204] A national infrastructure would support searches and deliver results appropriately compiled and processed to protect privacy. Locating all of a patient's data, wherever stored, would work rather the way an Internet search engine works today. Until such a solution is implemented, LHRs and LPHD will continue to require labor-intensive services. The hope, eventually, is to replace some of the human services with more capital-intensive infrastructure services.

The problem, always, has been how to mobilize the needed capital investment. P-CAST acknowledges that federal leadership would be required to create the needed infrastructure, since "market forces are unlikely to generate appropriate incentives for the necessary coordination to occur spontaneously."[205] This view is far more pessimistic than the view, expressed by Hall and Schulman, that altering patient's entitlements in their health data "would stimulate market development of interconnected electronic medical records (I-EMRs).[206] The problem with clarifying ownership in patients' health data is that it is a supply-side solution—and this remains true whether ownership is clarified in favor of patients (as in Hall and Schulman's proposal) or the public (as in Rodwin's). In contrast, health information infrastructure exhibits problems both on the supply side and on the demand side. An example is the P-CAST proposal just described: It could reduce delays in supplying LHRs and LPHD, but the market may not value the incremental

---

202 P-CAST REPORT, *supra* note 192, at 4.
203 *Id*. at 41.
204 *Id*.
205 *Id*. at 4.
206 Hall, *supra* note 6, at 636.

speed, since many health data applications are not latency-sensitive. Researchers who can afford to wait for a good dataset may not be willing to pay more for a good dataset delivered sooner. That is a demand-side issue.

There are others. The price users would be willing to pay for health data resources may not reflect the true value of those resources because so many uses of health data (such as research and public health activities) themselves produce public and nonmarket goods.[207] In this situation, data users are unable to appropriate the full value their activities create and therefore will not reflect it in the price they pay for data resources.[208] Frischman has noted that market failure for infrastructure is more complex than supply-side analysis suggests.[209] "For both traditional and nontraditional infrastructure resources, analysts emphasize supply-side issues … and assume that the market mechanism will best generate and process demand information."[210] Data propertization "solutions" assume that if encounter-level patient data were simply assigned to the right owner, the market would be able to figure out the right price to pay for useful data resources such as LHR and LPHD, and this price would cover the cost of necessary infrastructure and services to make those resources. This is not a safe assumption.

IV. THE HITECH ACT'S STRATEGY FOR PROMOTING INFRASTRUCTURE DEVELOPMENT

The HITECH Act, which Congress passed as part of 2009 economic stimulus legislation,[211] is accused of having done little to promote interconnection of health information systems. Hall notes that, while providing funds to help providers and physicians install electronic health records systems, "the economic stimulus act contains no legal requirement that funded systems actually interconnect to form a consolidated medical record for each patient."[212] Rodwin, in discussing the goal of "sharing of patient data for research and public uses" notes that "HITECH does not appear to authorize creating regulations that can achieve that goal."[213]

It is true that the HITECH Act does not expressly require interconnection of data systems or sharing of data, but it did something arguably more important: It clarified pricing of the infrastructure services that are required to convert raw patient data into valuable data resources for research and public health, and it authorized data-holders to conduct commercial transactions for sale of those services[214] when supplying data for public health and research. In doing so, it set the foundation for a commercial market in infrastructure services and supplied the heretofore-missing mechanism for financing private-sector development of health information infrastructure.

---

207 *See* Frischman, *supra* note 67, at 966-967 (defining and comparing public and nonmarket goods).

208 *Id.* at 968 (noting, "Infrastructure uses that produce public goods and nonmarket goods suffer valuation problems because they generally do not fully measure or appropriate the (potential) benefits of the outputs they produce and consequently do not accurately represent actual social demand for the infrastructure resource.").

209 *Id*. at 930.

210 *Id*.

211 American Recovery and Reinvestment Act of 2009 (ARRA), Pub. L. No. 111-5, 123 Stat. 115 (Feb. 17, 2009).

212 Hall, *supra* note 6, at 635.

213 Rodwin, *supra* note 6, at 595.

214 *See* discussion *infra* this section.

A. The Regulated Price of Infrastructure Services

At first glance, the HITECH Act purports to restrict sales of health data.[215] It states a general rule that it is unlawful for HIPAA-covered entities and their business associates to exchange a person's protected health information for direct or indirect remuneration—in other words, to sell data—unless the person has authorized the transaction.[216] However, this restriction is tempered by a list of exceptions.[217] One of the exceptions lets entities like insurers, healthcare providers, and academic medical centers that supply data to researchers pursuant to a HIPAA waiver—that is, without the consent of the people the data describe—to charge a price that "reflects the costs of preparation and transmittal of the data".[218] Unless the data supplier wishes to charge a price higher than this cost-based fee, the individuals' permission is not required.[219] When data are supplied for public health uses, data-holders also can charge a fee, and Congress chose not to impose a cost-based cap on such fees.[220]

It may, at first, seem wrongheaded for data-holders to charge higher fees to supply public health uses of data, which traditionally have been viewed as having a greater social value than research.[221] Yet this policy makes sense if you assume that, at present, the data supply is infrastructure-constrained. Under such conditions, a higher fee would help support investment in needed systems[222] to resolve the constraint, thus promoting wider availability of data for use by public health agencies. By letting them pay more than researchers can pay for data provisioning, Congress was helping ensure adequate flows of data to important public health uses. Later, when the United States has completed installation of its basic health information infrastructure, it may make sense to cap the fees for public health uses. The HITECH act allows the Secretary of HHS to impose a cost-based cap at a later time[223] based on an evaluation of how it may affect availability of data.[224] OCR is evaluating now whether its cost-based fee structure also should apply to data supplied to public health users.[225]

Data supplied to researchers pursuant to a waiver would be subject to the cost-based cap on fees. In common parlance, "cost-based" means "at cost," so this does not at first sound like a promising pricing structure to spur investment in interconnected data systems. Its potential becomes clear only when the phrase "reasonable, cost-based fee" is read together with a large body of other infrastructure regulatory precedents. Health information systems are infrastructure[226] and the HITECH Act's cost-based fee for data preparation and transmission echoes cost-of-service pricing traditionally used in other American infrastructure industries.

---

215 42 U.S.C. § 17935(d).

216 *Id*. § 17935(d)(1).

217 *Id*. § 17935(d)(2).

218 *Id*. § 17935(d)(2)(B).

219 *Id.* (allowing sales priced at the cost-based fee to move under the Privacy Rule's waiver provision at 45 C.F.R. § 164.512(i) which allows disclosure to researchers without individual authorization).

220 75 *Fed. Reg.* at 40921 (proposing a new regulation at 45 C.F.R. § 164.508(a)(4)(ii)(A)).

221 *See, e.g.,* LAWRENCE O. GOSTIN, PUBLIC HEALTH LAW, 2ND ED. 47 (2008) (discussing the high value traditionally accorded to public health activities).

222 CHARLES F. PHILLIPS, JR., THE REGULATION OF PUBLIC UTILITIES, 172 (1993) (noting the necessity of adequate earnings to support development and expansion of the industry).

223 42 USC § 17935(d)(3)(B).

224 42 USC § 17935(d)(3)(A).

225 75 *Fed. Reg.* at 40891.

226 *See* JOSÉ A. GÓMEZ-IBÁÑEZ, REGULATING INFRASTRUCTURE 4 (2003) (defining infrastructure as "networks that distribute products or services over geographical space").

Historically, many infrastructure industries exhibited natural monopoly characteristics or other structural problems that made it unwise to let prices be set by market forces.[227] These concerns supplied the rationale for imposing cost-based pricing schemes.[228] Cost-of-service pricing was widely used in American infrastructure regulation dating back to the Interstate Commerce Act of 1887[229] which regulated railroads. Congress subsequently imposed it on the interstate shipping,[230] stockyard,[231] telephone,[232] telegraph,[233] trucking,[234] electricity,[235] natural gas,[236] and aviation[237] industries.[238]

The words "reasonable" and "cost-based" have well-developed meanings in U.S. infrastructure regulation. These meanings were shaped by more than a century of Supreme Court cases examining cost-based pricing structures in other infrastructure industries.[239] Under these precedents, the reasonable, cost-based fee for data preparation and transmittal must—as a matter of Constitutional law—let data-holders, when responding to requests for data, recover: (1) their variable and fixed operating costs of responding to requests for data, (2) capital costs of the information systems used in responding to the requests, and (3) a reasonable profit margin.[240]

---

227 Joseph D. Kearney & Thomas W. Merrill, *The Great Transformation of Regulated Industries Law*, 98 COLUM. L. REV. 1323, 1334 (1998). *See also* GÓMEZ-IBÁÑEZ, *supra* note 226, at 4-6 (2003) (discussing rationales for infrastructure regulation); PHILLIPS, *supra* note 222, at 51-60 (discussing natural monopoly characteristics and structural issues that may call for price regulation); Hank Intven, Jeremy Oliver & Edgardo Sepulveda, *Module 1-Overview of Telecommunications Regulation*, in THE WORLD BANK INFORMATION FOR DEVELOPMENT PROGRAM (*INFO*DEV), TELECOMMUNICATIONS REGULATION HANDBOOK § 1.1.1, box 1 – 1 (Hank Intven, ed., 2000) [hereinafter, "*INFO*DEV, TELECOMMUNICATIONS REGULATION HANDBOOK"] (listing as an objective of regulation of prevent abuses of market power such as anticompetitive behavior and excess pricing in situations where markets do not exist) and Hank Intven, Jeremy Oliver, Edgardo Sepulveda, *Module 5 – Competition Policy*, *id*. at § 5.2.2 – 5.2.4 (discussing specific market imperfections common in infrastructure industries such as telecommunications).

228 PHILLIPS, *supra* note 222, at 182-83; GÓMEZ-IBÁÑEZ, *supra* note 226, at 5-6.

229 Interstate Commerce Act, ch. 104, 24 Stat. 379 (1887) (codified as amended in scattered sections of 49 U.S.C. app.).

230 Shipping Act of 1916, ch. 451, 39 Stat. 728, 733–35 (1916) (codified as amended at scattered sections of 46 U.S.C. app.).

231 Packers and Stockyards Act of 1921, ch. 64, 42 Stat. 159 (codified as amended at 7 U.S.C. §§ 181–229b (2006)).

232 Communications Act of 1934, ch. 652, 48 Stat. 1064 (codified as amended at 47 U.S.C.A. §§ 151–614 (West 2001 & Supp. 2008)).

233 *Id.*

234 Motor Carrier Act of 1935, ch. 498, 49 Stat. 543 (codified as amended in scattered sections of 49 U.S.C.).

235 Act of 1935, ch. 687, 49 Stat. 838 (codified as amended at scattered sections of 16 U.S.C.).

236 Natural Gas Act of 1938, ch. 556, 52 Stat. 821 (codified as amended at 15 U.S.C. §§ 717–717w (2006)).

237 Civil Aeronautics Act of 1938, ch. 601, 52 Stat. 973 (codified as amended and before repeal at scattered sections of 49 U.S.C.).

238 Kearney & Merrill, *supra* note 227, at 1333-34.

239 *See* Barbara J. Evans, RIN 0991-AB57: Modifications to the HIPAA Privacy, Security, and Enforcement Rules Under the Health Information Technology for Economic and Clinical Health Act 4-12 (Docket No. HHS-OCR-2010-0016, Sept. 10, 2010), http://www.regulations.gov/#!documentDetail;D=HHS-OCR-2010-0016-0086.1 (discussing design of the cost-based fee for preparation and transmittal of data authorized in section 13405(d) of the HITECH Act, reviewing cases in which the U.S. Supreme Court ruled on the constitutionality of cost-of-service fee structures in other infrastructure regulatory contexts, and demonstrating that, to be constitutional, the reasonable, cost-based fee for data provisioning must be set at a level sufficient to cover variable costs, include an allowance for fixed operating costs and capital costs, and provide a fair rate of return on invested capital).

240 *Id*. *See also* PHILLIPS, *supra* note 222 (providing a comprehensive review of judicial decisions affecting cost-of-service rates for infrastructure services in several industries).

The HITECH Act's cost-based fee structure, if implemented in accordance with these precedents, would foster creation of a commercial market in the infrastructure services that are needed to convert encounter-level patient data into valuable data resources for research and public health.

In July, 2010, The Office for Civil Rights (OCR) within the U.S. Department of Health and Human Services (HHS) proposed to implement the fee by amending the HIPAA Privacy Rule.[241] The OCR's proposed regulation tracks the statute closely and would let data be sold for use in research under a HIPAA waiver[242] so long as the entity supplying the data receives only "a reasonable cost-based fee to cover the cost to prepare and transmit" the data.[243] The OCR sought public comments on how, precisely, it should define the cost-based fee: which cost items should and should not be included?[244] In defining the fee, the OCR is not writing on a blank slate. It will be bound to follow precedents from other infrastructure industries. Should it fail to do so, litigation can be expected to follow, and the precedents strongly favor data-holders' claims to receive full recovery of their operating and capital costs, plus a reasonable profit margin.

To be clear, the HITECH Act does not "monetize[e] medical information."[245] The cost-based fee is not actually a price for data; it is a price for services. Insurers, healthcare providers, and other entities that operate health databases often do not "own" the data[246] and hence are not in a position to sell it. The thing data-holders own is their health information infrastructure: computer systems, software, communications equipment in which they have invested to support their regular lines of business. The proposed fee is to cover services that the data-holders and their skilled personnel provide with the aid of that infrastructure. The HITECH Act refers to these services as preparation and transmittal of data.[247] Colloquially, the term "data provisioning" is sometimes used to describe services, such as these, that make data available to users. With the aid of an insurance database, for example, it is possible to sift through large volumes of data, select information that meets a researcher's specifications, and process it for transmission to the researcher. The fee described in the HITECH Act[248] is for these sorts of infrastructure services. Technically speaking, the data are supplied at no charge and the fee is for services provided in responding to the data request.

Cost-of-service rates were common in U.S. infrastructure regulation until late in the 20th century, when they were partially replaced by targeted market-based reforms.[249] These reforms sought, whenever competitive conditions allowed, to let market forces play a greater role in setting the price of infrastructure services.[250] The modern critique of cost-of-service regulation focuses on its potential to be inefficient and cumbersome to administer.[251] This critique emerged

---

241 75 *Fed. Reg.* at 40921 (proposing a regulation to be codified at 45 C.F.R. § 164.508(a)(4)(ii)(B)).

242 45 C.F.R. § 164.512(i).

243 75 *Fed. Reg.* at 40921 (proposing a regulation to be codified at 45 C.F.R. § 164.508(a)(4)(ii)(B)).

244 75 *Fed. Reg.* at 40891 (seeking public comment on what should be included in the cost-based fee).

245 *See* Hall, *supra* note 6, at 651 (noting, "Law either prohibits monetizing medical information, or does not clearly permit this" and proposing to allow patients to sell rights to their data).

246 *See supra* notes 12-17 and accompanying text.

247 42 U.S.C. § 17935(d)(2)(B).

248 *Id.*

249 *See* Jim Chen, *The Nature of the Public Utility: Infrastructure, the Market, and the Law*, 98 NW. U. L. REV. 1617, 1618 (2004) (reviewing GÓMEZ-IBÁÑEZ, *supra* note 226).

250 Kearney & Merrill, *supra* note 227, at 1333–40.

251 *See, e.g.*, Chen, *supra* note 249, at 1631 (noting the public utility regulation has been criticized as raising questions of "indeterminacy and inefficiency").

late in the 20th century when the major policy challenge was to optimize use of existing infrastructures, as opposed to getting new infrastructures financed and built.[252] Chen notes that traditional cost-of-service infrastructure regulation actually may be the more efficient approach under economic conditions that existed earlier in the 20th century.[253] At that time, policymakers' central challenge was to build new infrastructures. That is the same challenge policymakers now face with respect to America's health information infrastructure: to get it built.

Congress's choice of cost-based pricing is a promising approach to the problem of financing health information infrastructure. Governmental intervention in markets is justified when barriers—for example, economic or legal—are blocking private-sector development of necessary infrastructure.[254] Various forms of intervention are possible, ranging from industry-specific regulation[255] to outright public ownership and operation of infrastructure.[256] The United States has rejected the latter option consistently throughout our nation's history. The U.S. is the only nation that maintained private ownership of its major infrastructure networks, such as pipelines and power grids, throughout the entire twentieth century.[257] It did, however, pervasively regulate these industries, including regulation of pricing. Other nations embraced public infrastructure ownership in varying degrees,[258] either from the very outset or by nationalizing privately owned infrastructures during the middle decades of the 20th-century.[259] Late the in the 20th century, there was a trend back to private infrastructure ownership with many nations implementing infrastructure privatization programs.[260]

The HITECH Act's data sales provisions are a traditional American approach to the problem of getting major, new infrastructure developed. Rather than have the government build big databases or otherwise own health information infrastructure, the HITECH Act presumes infrastructure will be developed, owned, and operated by the private sector subject to regulated pricing of infrastructure services. Congress reached for a pricing formula that has successfully financed private-sector development of large infrastructures for 150 years.

B. Where Things Stand

To summarize, this is the state of affairs after passage of the HITECH Act: Encounter-level patient data are an input that can be transformed into high-valued data resources, LHR and LPHD, for use in clinical care, research, and public health activities. Making these high-valued data resources also requires inputs of human and infrastructure services (data provisioning services). In theory, it is possible to produce LHRs for use in clinical care under a patient-controlled system. Such a system would subject all transfers of encounter-level patient data to consensual ordering (that is, permission of the patients whose data are involved). There are major

---

252 *Cf. id.* at 1620–21 (discussing the changes in infrastructure priorities from the nineteenth to twentieth centuries).

253 *Id.* at 1633, 1650.

254 *See* PHILLIPS, *supra* note 222, at 172–73; *see also* GÓMEZ-IBÁÑEZ, *supra* note 226, at 20–21; Chen, *supra* note 249, at 1624–28 (reviewing Gómez-Ibáñez's discussion of government regulation of infrastructure operation).

255 Chen, *supra* note 249, at 1628.

256 *See id.* at 1629 (citing GÓMEZ-IBÁÑEZ, *supra* note 226, at 13); Daniela Klingebiel & Jeff Ruster, *Why Infrastructure Financing Facilities Often Fall Short of Their Objectives* 7 (World Bank Policy Research Working Paper, No. 2358, 2000).

257 GÓMEZ-IBÁÑEZ, *supra* note 226, at 2; *see also* Chen, *supra* note 249, at 1632 (citing STEVEN BREYER, REGULATION AND ITS REFORM, 181–83 (1982)).

258 *See* Chen, *supra* note 249, at 1634.

259 GÓMEZ-IBÁÑEZ, *supra* note 226, at 2.

260 Klingebiel & Ruster, *supra* note 256, at 7.

limitations to such a system, however: Because of consent bias, the system cannot supply unbiased LPHD for use in research and public health projects. Secondary research and public health uses thus cannot be counted on to cross-subsidize the costs of developing patient-controlled LHRs. Unless the costs of developing patient-controlled LHRs are justified by the value they create in clinical care, the system may not be financially viable.

Creating high-valued data resources for research and public health applications requires a framework of nonconsensual access to encounter-level patient data. The HIPAA Privacy Rule and the Common Rule both allow nonconsensual access to patients' data for public health and research uses. If patients owned their encounter-level data, nonconsensual access for these uses still would be possible through exercise of the police and eminent domain powers. Nonconsensual access to patient data is necessary, but not sufficient, to ensure an adequate supply of high-valued data resources for research and public health. Two groups of entities potentially can block production of LHR and LPHD: (1) patients, to the extent they are able to block access to encounter-level data held by various data-holders, and (2) data-holders, to the extent they are able to block access to the services for locating relevant data and putting it into a format that can be combined with data from other data-holders. This latter bottleneck is the harder one to resolve.

The state's police and eminent domain powers only allow nonconsensual transfers of *property*; there is no similar mechanism that lets the government require nonconsensual provision of *services*. Forced provision of services would amount to involuntary servitude. The government generally obtains services consensually, via paid contracts for services or by requiring services in exchange for participation in desirable programs (for example, by requiring hospitals to report data as a condition of their eligibility to receive Medicare payments). The prospective provision of services is inherently consensual under our system of law. Accordingly, the HIPAA Privacy Rule and Common Rule do not allow nonconsensual access to data provisioning services. Waivers only *permit* data-holders to disclose data but do not require them to do so. Under HIPAA, a waiver can be approved by the would-be data user, but such a waiver is not a "call option" on data: the data-holder may supply data pursuant to the waiver but cannot be required to do so. This is fair: data-holders do not have unlimited capacity to supply services and need discretion to refuse. Nonconsensual access to data is possible whether under a property regime or the HIPAA/Common Rule regulatory regime. Nonconsensual access to services is not possible under either regime.

The HITECH Act acknowledged that access to data provisioning services is inherently consensual. It authorizes a pricing structure that, if implemented properly, will create incentives for data-holders and other potential service providers to "come to the market"—that is, to make infrastructure services available within their existing capacities and to invest in capacity expansion.

VI. WHAT STILL NEEDS TO BE DONE

The HITECH Act's pricing provisions may improve the situation, but all is not well. There remains a widely-shared perception that the existing framework of regulation under the HIPAA Privacy Rule and Common Rule is blocking socially beneficial uses of data while still under-protecting individual privacy.[261] The HIPAA Privacy Rule and Common Rule evolved

[261] *See supra* notes 4, 5.

over a 28-year period that began when the National Research Act of 1974[262] called for formation of a National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research (National Commission). The period ended in 2002 when the HIPAA Privacy Rule was promulgated in its current form[263] The Privacy Rule borrowed key concepts, such as the use of private review boards to approve waivers, from the Common Rule. The Common Rule was designed primarily with a view to protecting subjects of interventional (clinical) research and behavioral research.[264] Oversight mechanisms that have performed fairly well in clinical research "are not easily exported and applied to the very different challenges of epidemiologic research."[265] In 2000, when HHS promulgated the first HIPAA Privacy Rule, it did so over the objection of multiple comments, including comments from several members of Congress, voicing these same concerns.[266] In the decade that followed, research with data[267] and tissues[268] has grown in importance, making the problems ever more visible.

Four things need to be done to adapt these regulations for application to informational research: (1) apply the Common Rule in a manner that restores the proper scope of the state's police power to use data in public health activities; (2) develop a workable "public use" requirement for nonconsensual use of data in research; (3) devise appropriate procedural protections for the waiver-granting process; and (4) delineate appropriate federal and state roles in oversight of privacy and data access. The discussion below focuses on these first two issues, heretofore neglected in the scholarly debate. The third item has been discussed elsewhere[269] and the fourth is too vast a problem to resolve within the space of this discussion.

A. Restoring the Proper Scope of the State's Police Power to Use Data to Promote Public Health

Regulatory practice under the Common Rule conceives the scope of the state's police power more narrowly than it is conceived in any other legal context. This can be traced to an "original sin" during design of the Common Rule: its framers failed to define public health actions or delineate when they should be exempt from the Common Rule's consent requirements. The National Commission was instructed to delineate the boundary between research and

---

262 National Research Act of 1974, Pub. L. 93-348 (July 12, 1974).

263 67 *Fed. Reg.* at 53192.

264 43 *Fed. Reg.* 56174 (discussing, in the National Commission's 1978, the various types of research for which human-subject protections were being designed).

265 Casarett *et al*., *supra* note 40, at 587.

266 65 *Fed. Reg.* at 82690-91 (responding to comments that the waiver provision, which appeared at section 164.512(j) of the proposed regulation and at 164.512(i) of the final regulation, was inadequate because it had been "modeled on the existing system of human subject protections" and that "the Common Rule's requirements may be suited for interventional research involving human subjects, but is [sic] ill suited to the archival and health services research typically performed using medical records without authorization.").

267 *See, e.g.*, AHRQ Fact Sheet, *supra* note 33; Fred D. Brenneman *et al*., *Outcomes Research in Surgery*, 23 WORLD J. SURGERY 1220 (1999).

268 *See, e.g.*, Barbara J. Evans & Eric M. Meslin, *Encouraging Translational Research Through Harmonization of FDA and Common Rule Informed Consent Requirements for Research with Banked Specimens*, 27 J. LEGAL MED. 119, 122 (2006); Rina Hakimian & David Korn, *Ownership and Use of Tissue Specimens for Research*, 292 JAMA 2500 (2004).

269 *See supra* note 91.

medical treatment.[270] There was no similar directive to consider the relationship between research and public health actions.

The Belmont Report[271]—which set the ethical principles embodied in the Common Rule—defined research as an activity that produces generalizable knowledge.[272] Using generalizability to mark the line between research and treatment worked well; it kept common "experimental" therapeutic practices, such as the off-label use of drugs in routine clinical care, from falling under the jurisdiction of the Common Rule.[273] This definition carried through into the Common Rule's definition of "human-subject research"[274] and HIPAA's definition of "research".[275] Generalizability has jurisdictional significance under the Common Rule: it delineates whether an activity is, or is not, "human-subjects research" that is regulated by the Common Rule (and thus subject to its informed consent requirements). It does not have similar significance under HIPAA which has status-based jurisdiction based on attributes of the data-holder.[276]

The problem, under the Common Rule, is that generalizability of results does not provide a good bright-line rule for determining whether public health actions should or should not require consent. For example, vaccinating people to control a smallpox epidemic is permissible even without their consent;[277] vaccinating people to see which of two vaccines works better is research that obviously should require consent. Nonconsensual vaccination is justified in the first case—not because it fails to produce generalizable results, but because the unvaccinated person poses a potential threat of contagion to others in the circumstances of an epidemic.

Since the Common Rule was implemented, there have been tortured efforts to draw a sensible line between "public health activities" (which do not require consent) and "public health research" (which does). Various analytical frameworks have been proposed; they consider multiple factors in addition to whether generalizable knowledge is being produced.[278] The fact

---

[270] U.S.. Dep't of Health, Educ., & Welfare, The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, Notice of Report for Public Comment, __ *Fed. Reg.* __ (April 18, 1979) [hereinafter, "Belmont Report"] (describing, in the preamble to Belmont Report, that the National Commission was directed to consider "the boundaries between biomedical and behavioral research and the accepted and routine practice of medicine").

[271] *Id*.

[272] *Id*.

[273] *Id*.

[274] *See* 45 C.F.R. 46.102(d) and (h) (defining "research" and "human subject").

[275] 45 C.F.R. 164.501.

[276] See HIPAA (defining covered entity. See revisions, extending coverage to business associates).

[277] Jacobson v. Massachusetts, 197 U.S. 11 (1905).

[278] *See, e.g.*, JAMES G. HODGE, JR., & LAWRENCE O. GOSTIN, COUNCIL OF STATE & TERRITORIAL EPIDEMIOLOGISTS, PUBLIC HEALTH PRACTICE VS. RESEARCH 7 (2004), *available at* http://www.cste.org/pdffiles/newpdffiles/CSTEPHResRptHodgeFinal.5.24.04.pdf; NAT'L INST. OF HEALTH, U.S. DEP'T OF HEALTH & HUMAN SERVS., PROTECTING PERSONAL HEALTH INFORMATION IN RESEARCH (2004), *available at* http://privacyruleandresearch.nih.gov/pdf/HIPAA_Booklet_4-14-2003.pdf; Paul J. Amoroso & John P. Middaugh, *Research vs. Public Health Practice: When Does a Study Require IRB Review?*, 36 PREVENTIVE MED. 250, 250–53 (2003); Ctrs. for Disease Control & Prevention, U.S. Dep't of Health & Human Servs., *HIPAA Privacy Rule and Public Health*, MORBIDITY & MORTALITY, WKLY. REP., Apr. 11, 2003, at 1 6-11, *available at* http://www.cdc.gov/mmwr/pdf/other/m2e4111.pdf; James G. Hodge, *An Enhanced Approach to Distinguishing Public Health Practice and Human Subjects Research*, 33 J.L. MED. & ETHICS 125, 127 (2005); Dixie E. Snider, Jr. & Donna F. Stroup, *Defining Research When it Comes to Public Health*, 112 PUB. HEALTH REP. 29 (1997); Ctrs. for Disease Control & Prevention, U.S. Dep't of Health and Human Servs.,

remains, however, that generalizability of results gives rise to a presumption that the activity is "research" that will require informed consent, and there is no clear, reproducible standard by which to overcome that presumption. Public health actions that produce generalizable knowledge, with minor exceptions,[279] require informed consent.

The activity-research distinction is extremely problematic as applied to public health uses of people's *data*, as opposed to public health actions that affect their bodies. Exercises of the state's police power can be enjoined only when they are illegitimate, as when the government acts beyond its constitutional powers or infringes a constitutional right.[280] Governmental touching of a person's data is not as constitutionally problematic as governmental touching of a person's body. The state can legitimately do things with people's data that it could not permissibly do with their bodies. Research is one of those things. Unconsented governmental research on people's bodies would implicate constitutional protections against bodily invasion. Unconsented research on data is not subject to those same constitutional boundaries. The activity-research distinction, when applied to uses of people's data, has the effect of drastically narrowing the scope of the state's police power in the area of public health.

The state, when legitimately exercising its police power, can require its citizens to enter nonconsensual transactions that benefit the public. *Any* legitimate exercise of the police power—including those that produce generalizable knowledge—can support the imposition of nonconsensual requirements on citizens. Nowhere, other than under the Common Rule, does law parse legitimate exercises of the police power into those that produce generalizable knowledge (and thus require consent) and those that do not. Indeed, actions that produce generalizable knowledge offer greater benefit to the public and, if anything, present a stronger case for nonconsensual access to data. Treating generalizability as grounds to require consent, as the Common Rule does, can yield the wrong answer in informational research: consent requirements are imposed *in inverse proportion* to the amount of public benefit the use will generate.

An example of this ongoing problem arose recently after Congress authorized development of a large health data network[281] for use in drug safety surveillance and various other activities that have the potential to produce generalizable knowledge. [282] Congress clearly has power to legislate to protect the public health.[283] It is almost inconceivable that modern courts would question Congress's determination that these activities offer public health benefits sufficient to warrant nonconsensual access to data[284] It thus seems singularly inappropriate for

---

Guidelines for Defining Public Health Research and Non-research (1999), http://www.cdc.gov/od/science/regs/hrpp/researchdefinition.htm; Office for Prot. from Research Risks, Office for Human Prots., OPRR Guidance on 45 C.F.R. § 46.101(b)(5): Exemption for Research and Demonstration Projects on Public Benefit and Service Programs, http://www.hhs.gov/ohrp/humansubjects/guidance/exmpt-pb.htm (last visited Nov. 14, 2008) [hereinafter OPRR Guidance].

279 *See* U.S. Dep't of Health and Human Servs., Guidelines for Defining Public health Research and Non-Research, *supra* note 279 (giving the example that it would be acceptable to make nonconsensual use health data of victims of a virus outbreak on a cruise ship to try to identify the cause, even though the knowledge gained is generalizable in the sense that it likely will benefit future cruise passengers).

280 *See* Merrill, *supra* note 53, at 65.

281 *See supra* notes 180-182 and accompanying text.

282 FDAAA § 905(a); 21 U.S.C. § 355(k)(3)(C)(i)(I)-(VI). *See* Evans, *supra* note 13, at 601-602 (discussing the purposes for which Congress authorized development of the Sentinel network)

283 *See* Parmet, *supra* note 60 (discussing the scope of the federal public health power).

284 *See* Merrill, *supra* note 53, at 63 (discussing, in a different context (takings), courts' "extreme deference" to legislative findings that an activity offers public benefit).

private IRBs to second-guess Congress's determination. However, some IRBs construe the Common Rule as empowering them to do so. To quiet the matter, it was necessary to obtain a determination from the Director of OHRP that the congressionally authorized data uses are public health activities lying outside the scope of the Common Rule.[285] Even then, a material number—about five percent--of IRBs refused to allow data access for one recent study.[286] To date, there has been a surprising lack of debate about whether it is appropriate for private IRBs to nullify legislative determinations of what is in the public interest.

The Privacy Rule properly frames legislatively authorized public health uses of data as legitimate exercises of the police power. It was designed specifically to regulate disclosure and use of data, as opposed to interventional activities. It contains an exemption allowing data-holders to make nonconsensual disclosures of data to a "public health authority that is authorized by law to collect or receive such information"[287] for purposes that are defined broadly enough to include studies that produce generalizable knowledge.[288] The data-holder does not need to conduct an IRB review or make any inquiry into the nature of the intended data use. It merely needs to verify that the person requesting the data is a public health official with legal authority to request the data,[289] and that the requested data are the minimum necessary to fulfill the public health purpose. [290] The data-holder is entitled to rely on the public health authority's representations that it has legal authority to make the request.[291]

---

285 Kristen Rosati, Barbara Evans & Deven McGraw, HIPAA and Common Rule Compliance in the Mini-Sentinel Pilot (Mini-Sentinel Coordinating Center, 2010), http://mini-sentinel.org/work_products/About_Us/HIPAA_and_CommonRuleCompliance_in_the_Mini-SentinelPilot.pdf, *see* Annex 1, *id*. at 10, Letter from Jerry Menikoff, Director, Office for Human Research Protections to Rachel Behrman, Acting Associate Director of Medical Policy, Center for Drug Evaluation and Research, U.S. Food & Drug. Admin., dated Jan. 9, 2010 (deeming Sentinel activities not to be regulated by the Common Rule).

286 *See* [forthcoming].

287 45 C.F.R. §164.512(b)(1)(i). *See also* 45 C.F.R. § 164.501 (defining public health authorities to include public agencies as well as entities acting under a contract with an agency).

288 *Id*. (including public health "investigations" as well as "interventions").

289 *See* 45 C.F.R. § 164.514(h)(2)(ii)(C) (allowing a covered entity, when making disclosure to a person acting on behalf of a public official, to rely on "a written statement on appropriate governmental letterhead that the person is acting under the government's authority or other evidence or documentation of the agency, such as a contract for services … that establishes that the person is acting on behalf of the public official"; 45 C.F.R. § 164.514(h)(2)(iii)(A) (permitting a covered entity to rely on the written statement of a public agency regarding the legal authority under which it is requesting PHI, or an oral statement if a written statement is impracticable). See also, 65 *Fed. Reg.* at 82547 (explaining, in the Preamble to the Privacy Rule, that the verification process can rely on "reasonable" documentation).

290 *See* 45 C.F.R. § 164.514(d)(3)(iii) ("A covered entity may rely, if such reliance is reasonable under the circumstances, on a requested disclosure as the minimum necessary for the stated purpose when: (A) Making disclosures to public officials that are permitted under § 164.512, if the public official represents that the information requested is the minimum necessary for the stated purpose." While §13405(b) of the Health Information Technology for Economic and Clinical Health Act (the HITECH Act), codified at 42 U.S.C. § 17935, contains a provision that requires covered entities to determine what is the minimum amount of PHI for a disclosure, recently proposed amendments to the HIPAA Privacy Rule to implement the HITECH Act do not modify a covered entity's ability to rely on minimum necessary representations by public officials. . . . . (See Notice of Proposed Rule Making, "Modifications to the HIPAA Privacy, Security, and Enforcement Rules under the [HITECH] Act," at http://www.ofr.gov/OFRUpload/OFRData/2010-16718_PI.pdf, scheduled for publication in the Federal Register on July 14th.)

291 *See supra* note 289.

This leaves two problems: In many data environments, the Common Rule and the HIPAA Privacy Rule both apply. The Common Rule, when it applies, continues to require an analysis of whether a public health use of data is an activity or research. This analysis penalizes public health uses that produce generalizable results, although those are precisely the uses of data most likely to yield public benefits and most deserving of access to data. OHRP should issue guidance--or, if unavoidable, amend the Common Rule—to conform it to HIPAA's more appropriate handling of public health uses of data. The second problem is that data-holders and IRBs, long accustomed to the Common Rule, have had difficulty appreciating that the HIPAA Privacy Rule rejected the Common Rule's approach to public health uses of data. IRBs and privacy boards administering public health disclosures under the Privacy Rule sometimes mistakenly continue to analyze whether a use is an activity or research. HIPAA does not require this, and such misunderstandings continue to hinder public health access to data.

B. Developing a Workable Doctrine of Public Use of Private Data

The existing pathways[292] for nonconsensual use of data under the Common Rule and HIPAA Privacy Rule were developed in an *ad hoc* manner to preserve specific uses of data that already had well-established histories before these regulations came into force. For example, health data had been widely used in research without consent in the decades before the Common Rule came into existence.[293] The regulations preserved preexisting uses without enunciating a coherent theory explaining why—and which—public uses of private data are appropriate. The waiver provisions of the HIPAA Privacy Rule and Common Rule lack a "public use" requirement—a criterion, similar to the one in eminent domain,[294] that requires nonconsensual research uses to serve a publicly beneficial purpose.[295] There is wide agreement among bioethicists that the "central ethical issue" [296] in health informational research is to ensure that the potential public benefits are sufficient to warrant the burden on the individual.[297] At every stage of the process that led to development of HIPAA and the Common Rule, advisory bodies

---

[292] *See* Evans, *supra* note 79, at 4 (summarizing pathways for nonconsensual use of data under the Common Rule and HIPAA Privacy Rule).

[293] *See* Dep't of Health, Educ., & Welfare, Office of the Secretary, Protection of Human Subjects: Institutional Review Boards: Report and Recommendations of the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 43 *Fed. Reg.* 56174, 56188 (Nov. 30, 1978) (noting that a survey of investigators, conducted as part of efforts to develop the Common Rule, found that the fact that a "study was based exclusively on the use of existing records" was commonly cited as a reason why consent was unnecessary—in other words, the prevailing norm before the Common Rule appears to have been not to require consent for research that relies on existing health records).

[294] *See* notes 61-67 *supra* and related text.

[295] Merrill, *supra* note 53, at 61.

[296] Casarett *et al*., *supra* note 40, at 597.

[297] *Id.* at 597 ("The central ethical issue in pharmacoepidemiologic research is deciding what kinds of projects will generate generalizable knowledge that is widely available and highly valued, and do this in a manner that protects individuals' right to privacy and confidentiality.") *See also* Peter D. Jacobson, *Medical Records and HIPAA: Is It Too Late to Protect Privacy?*, 86 MINN. L. REV. 1497, 1498 (2002) (arguing that the nub of the problem in health information privacy is to determine which public health objectives are sufficiently important to override the individual's interest in nondisclosure). See also NATIONAL BIOETHICS ADVISORY COMMISSION, 1 ETHICAL AND POLICY ISSUES IN RESEARCH INVOLVING HUMAN PARTICIPANTS (August, 2001), http://bioethics.georgetown.edu/nbac/human/overvol1.pdf (recognizing the need for nonconsensual data use in some circumstances and including, as a criterion, that an IRB determine that "the benefits from the knowledge to be gained from the research study outweigh any dignitary harm associated with not seeking informed consent").

that pondered nonconsensual research use of data called for a utilitarian balancing of public and private interests. It is worth tracing this history because it came to an anomalous result: the waiver provisions of the HIPAA Privacy Rule and Common Rule, as finally promulgated, have no criteria requiring such a balancing. The Common Rule and the HIPAA Privacy Rule provide the public with no assurance that unconsented uses of data will serve any socially beneficial purpose at all.

How the public use requirement got lost. The earliest precursor of the Common Rule was a set of 1974 regulations[298] that required informed consent and IRB review of research. There was no provision letting IRBs waive the consent requirement. The National Commission's recommendations about human-subject protections, published in 1978,[299] focused primarily on interventional and behavioral research. It discussed waiving or altering consent, but not with respect to the use of preexisting stores of data.[300] The report separately addressed research that relies on existing documents, records, or tissue specimens and stated several principles: If the data subjects are not identified or identifiable, such research need not be viewed as human-subjects research at all,[301] and consent requirements should not apply. Even when the data are identified, informed consent still may not be necessary, provided certain conditions are met.[302] These conditions included a public use requirement: an IRB must determine that "the importance of the research justifies such invasion of the subjects' privacy."[303]

HEW commenced proceedings[304] in 1979 to incorporate the National Commission's recommendations into its existing regulations. The proposed regulation did not include a waiver provision. HEW explained that it was instead considering whether certain types of behavioral research and research with data should be exempt from the regulations altogether (that is, not subject to a consent requirement at all).[305] HEW sought comments on how to handle research with data. What is striking is that the unconsented use of data was, at that time, a matter of considerable indifference. Fewer than 20 commenters discussed the proposed exemption for studies with existing data,[306] whereas other issues in the proceeding drew over 500 comments.[307] Most of those who commented favored exempting research uses of data from consent

---

298 U.S. Dep't of Health, Educ. & Welfare, Title 45, Subtitle A, Part 46 Protection of Human Subjects, 39 *Fed. Reg.* 18914 (May 30, 1974).

299 See *supra* note 40.

300 43 *Fed. Reg.* at 56180-81 (discussing consent waivers for certain types of behavioral research that study people who are unaware why they are being observed).

301 *Id*. at 56181.

302 *Id*. at 56179-80.

303 *Id*. at 56179. *See also id*. at 56181 (reporting findings of a Privacy Protection Study Commission, under the auspices of the National Commission, which elaborated this balancing requirement more specifically: "medical records can legitimately be used for biomedical or epidemiological research, without the individual's explicit authorization," provided that the medical care provider (who in all likelihood would have been the data-holder in that era of paper records) determines "that the importance of the research or statistical purpose for which any use or disclosure is to be made is such as to warrant the risk to the individual from additional exposure of the record or information contained therein" and provided that an IRB ensures this condition has been met).

304 U.S. Dep't of Health, Educ. & Welfare, Proposed Regulations Amending Basic HEW Policy for Protection of Human Research Subjects, 44 *Fed. Reg*. 47688 (Aug. 14, 1979) (to be codified at 45 C.F.R. part 46).

305 *Id*. at 47691.

306 U.S. Dep't of Health & Human Servs., Office of the Secretary, Final Regulations Amending Basic HHS Policy for the Protection of Human Research Subjects, 46 *Fed. Reg.* 8366, 8372 (Jan. 26, 1981) (to be codified at 45 C.F.R. part 46)

307 *Id*. at 8368.

requirements altogether. The final rule exempted research in which the investigator records data a de-identified manner.[308] This exemption still exists in the modern Common Rule.[309]

What was not addressed was whether consent could be waived for research that requires access to identified or identifiable data. This type of research later gained importance as post-1980 advances in information technology made it possible to link patients' records from multiple sources to form LHRs.[310] Longitudinal linkage of this type requires at least some access to identifying information to ensure that the various records being linked all pertain to the same patient.[311] The National Commission, in its 1978 report, had called for a mechanism to allow unconsented research access to identified data and records.[312] HEW and its successor, HHS, did not address this recommendation in their 1979-81 rulemaking.

The final regulation promulgated in 1981 did, however, insert a waiver provision[313] identical to the one that still exists in the Common Rule.[314] In explaining why it had, so late in the regulatory proceedings, inserted this provision, HHS made no reference to nonconsensual data use. Rather, the waiver provision was a response to an altogether different problem: research into the optimal design of federal benefit programs.[315] This explains why the Common Rule's waiver provision contains no public use requirement for nonconsensual data uses. The waiver provision, when initially implemented, was not intended for use in approving nonconsensual uses of data, so it did not incorporate the balancing test the National Commission had recommended. [316] Later, when the waiver provision was pressed into service for approving nonconsensual data uses, nobody thought to go back and amend the waiver criteria for this new purpose.

The HIPAA waiver provision presents a different story. The HIPAA Privacy Rule, though it has been criticized, was the product of a thoughtful and well-researched rulemaking process.[317] When developing its proposed regulation, HHS understood that waiving consent for research use of data raises issues that would not be adequately addressed by simply copying the waiver criteria of the Common Rule.[318] Instead, HHS started from scratch and proposed a whole new set of waiver criteria. These included a requirement that an IRB or privacy Board make a determination that "the research is of sufficient importance so as to outweigh the intrusion on the

---

308 *Id*. at 8387.

309 45 C.F.R. 46.101(b)(4).

310 See *supra* Part III.

311 See discussion *supra* notes 168-70 and accompanying text.

312 43 Fed Reg at 56179-80.

313 46 *Fed. Reg.* at 8390.

314 45 C.F.R. 46.116(d).

315 46 *Fed. Reg.* at 8383. HHS was responding to Crane v. Mathews, 417 F. Supp 532 (1976), which had held that IRB review should have applied to certain randomized studies (which varied Medicaid benefits to observe impacts on beneficiaries' consumption of health care). HEW had responded hastily with a strained interpretation that attempted to place such studies outside the scope of its regulations. See U.S. Dep't of Health, Educ. & Welfare, Secretary's Interpretation of "Subject at Risk," 41 *Fed. Reg.* 26572 (Jun. 28, 1976). The issue continued to simmer and, as HHS promulgated the final revised regulations in 1981, it tried a different solution: HHS admitted that such research should be subject to IRB review but added a waiver provision to let informed consent be waived. 46 *Fed. Reg.* at 8383.

316 43 *Fed. Reg.* at 56181

317 65 *Fed. Reg.* 82462 (discussing, in the preamble to the initial HIPAA Privacy Rule, the rationale for its various provisions).

318 65 *Fed. Reg.* at 82697 (noting that the Common Rule's waiver criteria were not explicitly directed at protecting the privacy interests that the HIPAA privacy rule protects).

privacy of the individual whose information is subject to the disclosure."[319] Unfortunately, this criterion drew "a large number" of adverse comments.[320] Some commenters warned that the criterion was subjective and would be inconsistently applied by IRBs; others criticized its reliance on conflicting value judgments as to whether research is important.[321] One unyielding privacy advocate declared that public purposes should never be able to override individual interests in a democratic society.[322]

Some commenters noted that IRBs already balance risks and benefits of research when applying 45 C.F.R. sec. 46. 111(a)(2), which is a criterion for IRB approval of *any* type of research whether consented or nonconsented.[323] This criterion requires that the risks of research be reasonable in relation to the anticipated benefits of the research (if any) to the individual and the importance of the knowledge that may reasonably be expected to result from the research. HHS accepted the commenters' suggestion to model the HIPAA waiver criterion on this provision, and this change was reflected in the December, 2000 version of the HIPAA Privacy Rule.[324] This change was wrongheaded. The criterion at sec. 46.111(a)(2) of the Common Rule is a minimum threshold for acceptability of research. Research that does not meet this criterion is considered so devoid of scientific merit that an IRB must not allow people to consent to it even if they wish to do so.[325] This is similar to the notion that some transactions—such as selling one's children—are so bad that a paternalistic state must step in and forbid the transaction, even if people want to do it. Adopting the criterion of minimal acceptability as the criterion for approving a waiver was nonsensical: in any situation where consent can be allowed, it can be waived. Any research that met the most minimal threshold of scientific non-odiousness qualified for a waiver under this criterion. This was not the sort of public use requirement the National Commission had proposed.

HHS had an opportunity to correct this error two years later, when a new administration asked HHS to revisit the HIPAA Privacy Rule.[326] Unfortunately, the correction took the form of jettisoning the troublesome balancing requirement altogether.[327] The currently effective HIPAA waiver provision, like its counterpart in the Common Rule, has no requirement that the proposed research offer any public benefit. These waiver provisions are functionally equivalent to a private delegation of takings power,[328] but it is a power devoid of any public use requirement. This is troubling when, as here, the power is coupled with a level of procedural informality[329] that offers little assurance against its abuse. Everybody is harmed: researchers' access to data is thwarted by the absence of a coherent doctrine for determining which uses warrant nonconsensual access, and data subjects perceive—rightly—that their "central ethical issue"[330] is being ignored.

---

319 *Id*. at 82698.
320 *Id*.
321 *Id*.
322 *Id*.
323 *Id*.
324 *Id*.
325 *See* 43 *Fed. Reg.* 56180 (discussing, in the National Commission's report, the notion that subjects should not be exposed to research that falls below a minimal threshold of scientific quality).
326 67 *Fed. Reg.* 53182
327 *Id*. at 53270 (to be codified at 45 C.F.R. 164.512(i)).
328 *See* discussion *supra* at Part II.
329 See *supra* note 91 and accompanying text.
330 Casarett *et al*., *supra* note 40, at 597.

Clarifying the concept of public use of data in research. Those who expressed concern, during the first HIPAA rulemaking, about IRBs' ability to balance public and private interests[331] may have had a point. Utilitarian balancing is fundamentally at odds with the autonomy-based bioethical principles these regulations were designed to uphold. The interests in the balance are incommensurable.[332] Miller has pointed out that even if research has high social value, and even if consent is logistically difficult or impossible to obtain, and even if a consent requirement may undercut the scientific validity of results, these facts "do not in themselves constitute valid ethical reasons for waiving a requirement of informed consent."[333]

The field of bioethics has drawn heavily on an atomistic concept of autonomy that portrays individuals as "self-reliant, self-governing, and fundamentally alone."[334] Tauber has remarked that foundational works of modern bioethics from the years 1954-70 fail to delineate how the principle of autonomy competes with other moral tenets.[335] After 1980, bioethicists began to explore alternative views of autonomy as not merely an "internal, psychological characteristic but also an external, or social" one,[336] with individuals achieving autonomy in cooperation rather than in isolation.[337] Alternatives to a consent-based model have been proposed[338] but none has addressed the practical mechanics of *how* to make decisions to allow nonconsensual use of data in service of broader public interests. "If the self is understood as a confluence of relationships and social obligations that are constitutive of identity; then autonomy may legitimately be subordinated to other moral principles that determine how the self is governed within a social context."[339] When, how, and *just how far* autonomy legitimately may be subordinated remains largely unstudied. The bioethics literature has not resolved what it means to respect autonomy in situations where binding, collective decisions must be made. Modern takings jurisprudence has been equally unable to solve this problem.[340]

Nonconsensual research use of data is a "muddle" [341] strikingly similar to the one that has afflicted regulatory takings jurisprudence[342] in the years since Penn Central Transportation Co. v. New York City.[343] In that case, the Supreme Court applied a utilitarian balancing of public and private interests to deny compensation to Penn Central when the city Landmark Commission

---

331 *Id*.

332 For examples of the analogous critique of balancing in other contexts, see T. Alexander Aleinikoff, *Constitutional Law in the Age of Balancing*, 96 YALE L.J. 943 (1987); Richard H. Pildes, *Avoiding Balancing: The Role of Exclusionary Reasons in Constitutional Law*, 45 HASTINGS L.J. 711 (1994); Jed Rubenfeld, *The First Amendment's Purpose*, 53 STAN. L. REV. 767 (2001).

333 Franklin G. Miller, *Research on Medical Records Without Informed Consent*, 36 J.L. MED. & ETHICS 560 (2008) (discussing but not necessarily endorsing this view).

334 ALFRED I. TAUBER, PATIENT AUTONOMY AND THE ETHICS OF RESPONSIBILITY 13 (2005).

335 *Id*. at 16.

336 *Id*. at 77, at 120 (citing (citing GRACE CLEMENT, CARE, AUTONOMY, AND JUSTICE: FEMINISM AND THE ETHIC OF CARE 22 (1996)).

337 *Id*. at 122.

338 IOM, PRIVACY REPORT, *supra* note 4, at 254-55 (reviewing and discussing several models).

339 TAUBER, *supra* note 334, at 85.

340 *See* Merrill, *supra* note 53, at 63-64 (lamenting the lack of clear standards for determining public use).

341 *See* Carol M. Rose, *Mahon Reconstructed: Why the Takings Issue is Still a Muddle*, 57 S. CAL. L. REV. 561 (1984); Louise A. Halper, *Why the Nuisance Know Can't Undo the Takings Muddle*, 28 IND. L. REV. 329 (1995).

342 *See* Claeys, *supra* note 49, at 1555 (noting, "modern regulatory takings law is widely recognized to be a 'muddle.').

343 438 U.S. 104 (1978).

restricted its ability to develop the airspace above Grand Central Station, even though the restriction inflicted a major financial loss on Penn Central for the public's benefit. The court applied a deferential "rational basis" review that presumed "regulation has high social value whenever it is 'reasonably related to the promotion of the general welfare.'"[344]

The waiver criteria, by abandoning even the attempt to perform utilitarian balancing, venture even farther, making an automatic presumption that research has high social value. This arguably may be the right decision: research generates positive externalities and it is hard to assess *a priori* which lines of research will ultimately pay off or how great the payout will be. It may be that it benefits society to encourage all research;[345] however, this is a decision that a society needs to make more deliberatively. Considering the history of how today's waiver criteria got to be the way they are, it is obvious no conscious decision ever was made. The current presumption that all lines of research have vast social value simply fell from the sky of rulemaking accidents.

For waivers to merit public trust, a workable public use criterion needs to be enunciated. The takings muddle suggests, by analogy, that there will be no easy solution. However, it suggests a number of possible approaches to explore.

1. ***Reject utilitarian balancing in favor of natural-rights analysis***. Nineteenth-century state courts analyzed takings cases under natural-rights principles that grounded property rights in personhood[346]—an approach that bears considerable resemblance to modern bioethical analysis that grounds privacy rights in autonomy. Claeys has argued rather persuasively that the old natural-rights analysis did a better job of drawing sensible lines than modern utilitarian balancing can do.[347] Of particular interest are cases where state actions force individuals to contribute positive externalities to the community (for example, laws requiring homeowners to install curbs at their own expense[348]). Such cases require courts to decide whether the action is a noncompensable exercise of police power or a compensated taking.[349] This line-drawing bears conceptual similarities to the problem of distinguishing public health uses from research uses of data. In the latter problem, compensation is not at stake;[350] what is at stake is whether the activity will be subject to the Common Rule's oversight requirements.

---

344 Claeys, *supra* note 49, at 1557 (quoting 438 U.S. at 131). See also, Merrill, *supra* note 53, at 63 - 65 (discussing judicial deference to legislative findings of public benefit).

345 *But see*, DANIEL CALLAHAN, WHAT PRICE BETTER HEALTH: HAZARDS OF THE RESEARCH IMPERATIVE (2006) (challenging the notion that medical research is inherently good and to be pursued without regard to the burdens it places on competing values).

346 Claeys, *supra* note 49, at 1577-86 (discussing 19th-century state courts' natural-rights analysis of eminent domain cases involving state actions to protect public health, safety, morals, and order or to abate private nuisances).

347 *See* Claeys, *supra* note 49 (comparing 19th-century takings cases that bore similarity to regulatory takings cases and comparing them to 20th century regulatory takings cases). See Id at 1556 (noting that "[t]akings law gets muddled only when it applies a certain kind of utilitarian property theory").

348 *See* Palmyra v. Morton, 25 Mo. 593, 594 (1857) (upholding a town ordinance requiring homeowners to curb and pave footpaths in front of their homes at their own expense).

349 *See* Merrill, *supra* note 53, at 65 (describing a continuum of consensual transactions, compensated takings, and uncompensated confiscation or interference with property rights under the police power).

350 See *supra* notes 71-73 and accompanying text (explaining why nonconsensual research uses of data would not be compensable even if data were patient-owned).

Natural-rights analysis held that owners are not entitled to takings compensation when they receive "implicit in-kind"[351] compensation—for example, when each homeowner who is forced to make improvements enjoys "reciprocity of advantage"[352] and benefits from the improvements others are forced to install.[353] This was cast as the state using its police powers to force a mutually advantageous exchange that would be hard for individuals to organize by themselves; each affected person gives something to, and gets something from, the others. When there was no reciprocity of advantage—that is, when the burdens of a measure to benefit the public are disproportionately visited on some members of the community[354]—the action was a taking and compensation was owed.

The notion of reciprocity of advantage survives in modern bioethical criteria for assessing whether a particular data use is a public health activity that can be conducted without informed consent. If benefits of a study will flow primarily to the people whose data are used, as opposed to being generalizable to other populations, this weighs in favor of a finding that the study is a public health activity.[355] The criterion of "benefits internal to the community" is simply "reciprocity of advantage" under a different name. Unfortunately, it is used in combination with other criteria—such as generalizability of results—that often muddy the waters. The 19th-century cases made reciprocity of advantage a central focus of analysis: Is the state singling out individuals to bear burdens for the benefit of others, or is the state forcing a mutually advantageous exchange?

A similar focus could help identify which nonconsensual uses of data are acceptable and warrant public trust even in an environment of strong respect for individual autonomy. Nonconsensual research use of data held in large regionally or nationally scaled data networks can be conceptualized as a mutually advantageous exchange. "At the conceptual limit, where one-hundred percent of the present and future drug-consuming 'community' is in the data set, benefits of studying the data are completely internal to that community"[356] and there is reciprocity of advantage. Each person gives something to, and gets something from, the community. In this light, research in very large data networks actually has stronger ethical justification than does

---

351 Claeys, *supra* note 49, at 1589 (citing RICHARD A. EPSTEIN, TAKINGS: PRIVATE PROPERTY AND THE POWER OF EMINENT DOMAIN 195 – 215 (1985) and Frank I. Michelman, *Property, Utility, and Fairness: Comments on the Ethical Foundations of "Just Compensation" Law*, 80 HARV. L. REV. 1165, 1225-26 (1967)).

352 *Id*. at 1587-89, 1619-21 (tracing the "reciprocity of advantage" or "common benefit of all" concepts in 19th- and early 20th-century state and federal cases and noting the 20th century trend to supplant natural-law analysis with utilitarian principles) and at 1633 (noting occasional references to reciprocity of advantage in modern Supreme Court cases but observing that modern applications have "diluted the principle so much that it is now meaningless").

353 *Id*. At 1557, 1589 (citing Paxson v. Sweet, 13 N.J.L. 196, 199 (1832)).

354 *See* Claeys, *supra* note 49, at 1570 (discussing the early case, Van Horne's Lessee v. Dorrance, *2 U.S. (2 Dall.) 304 (C.C.D. Pa 1795)* that enunciated the notion that there is a taking when governmental action lays "a burden upon an individual, which ought to be sustained by society at large*" 2 U.S. at 310). See also* Armstrong v. United States, *364 U.S. 40 (1960)* (acknowledging the existence of a taken when governmental action forces "some people alone to bear public burdens which, in all fairness and justice, should be borne by the public as a whole." *id*. at 49).

355 Hodge & Gostin, *supra* note 278.

356 Evans, *supra* note 13, at 616.

research with smaller datasets that force "some people alone to bear public burdens which, in all fairness and justice, should be borne by the public as a whole."[357]

2. ***Focus not on how decisions should be made, but by whom***. Deciding which lines of research offer substantial social benefit requires a global perspective that local IRBs do not possess. A centralized, national oversight body or a legislature is better positioned to assess which lines of research warrant nonconsensual data use. Establishing a publicly accountable body to identify general categories of research that offer public benefit would be one possible approach. Patient advocacy groups could petition it to allow data access for research into their "pet" diseases, much as they lobby Congress for research funding for specific diseases today.[358] When Congress has authorized specific lines of health informational research, as it did in FDAAA[359] and in the comparative effectiveness provisions of the Patient Protection and Affordable Care Act of 2010 (PPACA),[360] this should be treated as a Blackstonian "consent of the people" to the research. IRBs should not be tasked with second-guessing determinations of public benefit that a duly elected legislature has already made. Through guidance, OHRP and OCR could use their enforcement discretion to create a safe harbor in which data-holders would be deemed to have complied with the regulations when they make data available for legislatively authorized research uses.
3. ***Develop rules of thumb for identifying suspect, "non-public" uses of data.*** Merrill[361] has suggested the approach of identifying attributes of takings that mark them as likely to be devoid of a valid public purpose. These "presumptively private" uses then can be singled out for a more skeptical review. This approach may be helpful in delineating publicly beneficial research uses of data: specifying what they are not may be easier than specifying what they are. The idea would be to develop a list of "red flags" that lower the presumption that a proposed research use of data offers public benefit. To take one example, there is not presently a health informational research registry that serves the same purpose as ClinicalTrials.gov,[362] where sponsors of clinical trials disclose information about their planned projects. It sometimes is alleged that academic and commercial researchers would be reluctant to disclose their planned informational research activities, since doing so would give away their corporate strategies and research ideas. Unwillingness to disclose research plans might be viewed as a "red flag" that signals a data use has a primarily private purpose (personal ambition, commercial interest, etc.). Private-purpose data uses still could go forward, but they would require

---

357 Armstrong v. United States, 364 U.S. 40, 49 (1960).

358 *See* REBECCA DRESSER, WHEN SCIENCE OFFERS SALVATION: PATIENT ADVOCACY & RESEARCH ETHICS (2001) (discussing the role of patient advocacy groups in influencing national research policy and allocation of resources to research).

359 FDAAA § 905(a); 21 U.S.C. § 355(k)(3)(C)(i)(I)-(VI). *See* Evans, *supra* note 13, at 601-602 (discussing the purposes for which Congress authorized development of the Sentinel network)

360 Pub L. 111-148, 124 Stat. 119 (Mar. 23, 2010). *See id*. at § 6301 (amending Title XI of the Social Security Act 42 U.S.C. 1301 et. Seq. by adding Part D—Comparative Effectiveness Research).

361 *See* Merrill, *supra* note 53, at 90-92 (identifying a need for heightened scrutiny of takings in which there is high subjective valuation of the taken property; there is a potential for secondary rent seeking, and where there has been an intentional or negligent bypass of a thick market).

362 *See* ClinicalTrials.gov, www.clinicaltrials.gov (noting, "ClinicalTrials.gov is a registry of federally and privately supported clinical trials conducted in the United States and around the world. ClinicalTrials.gov gives you information about a trial's purpose, who may participate, locations, and phone numbers for more details.").

informed consent. Persons wishing to use the public's data should be willing to disclose what they intend to do with it. Other red flags could be identified.

A major concern is whether the Common Rule will need to be amended to address the special problems of informational research. Everybody hopes not. The Common Rule is like the U.S. Constitution in one important respect: both documents are very hard to amend. Amending the Common Rule requires the 18 federal agencies that implement it to promulgate 15 new regulations,[363] all converging on identical amendments if the Common Rule's "common-ness" is to be preserved. When legal text is hard to change, laws can be kept modern by interpreting them; old words are deemed to have new meanings that accommodate the change in circumstances. The Office for Human Research Protections (OHRP), which implements the Common Rule, has previously made use of interpretive guidance to refresh the regulation for research with tissues and data.[364] It is not a foregone conclusion that amendments to the Common Rule will be required.

The use of guidance –by OHRP for the Common Rule, and by HHS's Office of Civil Rights (OCR) for the HIPAA Privacy Rule—offers a promising pathway for installing protective constraints around the use of waivers. For example, guidance could be used to clarify the regulations' criteria for granting a waiver. OHRP and OCR could list specific privacy and data security protections that prospective data users must have in place, before an IRB may deem a research project to pose "minimal risk" to the data subjects. The use of waivers could be narrowed by construing what it means for it to be "impracticable" to obtain consent or "impracticable" to do the research without nonconsensual access to data. IRBs might be instructed, through guidance, to find the requisite "impracticability" only if the research requires more than 100,000 (or ten million) records and if researchers have shown why consensual assembly of data would unacceptably bias their results. Such restrictions would be analogous to land development statutes that delegate takings power to a private body, but restrict the power so that it only can be used to assemble large parcels of land for major redevelopment projects.[365] Through guidance, OHRP and OCR also could create a safe harbor in which IRBs would be

---

[363] A total of 18 federal agencies follow the Common Rule, which has been codified in the federal regulations at 15 locations applying to 16 federal agencies. In addition, the Central Intelligence Agency follows the Common Rule, Exec. Order No. 12,333, 46 *Fed. Reg.* 59,941, 59,952 (Dec. 4, 1981), and the Social Security Administration participates in the Common Rule, Social Security Independence and Program Improvement Act of 1994, Pub. L. No. 103-296, §106(b), 108 Stat. 1464, 1476 (1994). The 15 regulations are: 45 C.F.R. pt. 46, subpt. A (HHS); 7 C.F.R. pt. 1c (2005) (Department of Agriculture); 10 C.F.R. pt. 745 (2005) (Department of Energy); 14 C.F.R. pt. 1230 (2005) (National Aeronautics and Space Administration); 15 C.F.R. pt. 27 (2005) (Department of Commerce); 16 C.F.R. pt. 1028 (2005) (Consumer Product Safety Commission); 22 C.F.R. pt. 225 (2005) (International Development Cooperation Agency; Agency for International Development); 24 C.F.R. pt. 60 (2005) (Department of Housing and Urban Development); 28 C.F.R. pt. 46 (2005) (Department of Justice); 32 C.F.R. pt. 219 (2005) (Department of Defense); 34 C.F.R. pt. 97 (2005) (Department of Education); 38 C.F.R. pt. 16 (2005) (Department of Veterans Affairs); 40 C.F.R. pt. 26 (2005) (Environmental Protection Agency); 45 C.F.R. pt. 690 (2005) (National Science Foundation); 49 C.F.R. pt. 111 (2004) (Department of Transportation). The Office of Science Technology Policy, because it does not conduct or sponsor research, has not codified the Common Rule even though it signed the Federal Policy that supplied the text of the Common Rule. The U.S. Food and Drug Administration does not implement the Common Rule, instead implementing its own framework of human-subject protections (21 C.F.R. pts. 50, 56) which, while similar to the Common Rule, differ in important respects that bear on the regulation of research with tissues and data. *See* Evans & Meslin, *supra* note 268, at 119 (discussing these differences).

[364] OHRP 2004 Guidance, *supra* note 78.

[365] Bell, *supra* note 50, at 570-71.

deemed to have complied with the regulation if they follow a set of "public-regarding norms"[366] designed to protect persons whose data are being transferred. These could include procedural protections as well as substantive criteria designed to channel waiver approvals toward publicly beneficial uses. Much could be done to enhance public trust in, and public accountability of, the waiver process without amending the regulations.

VI. CONCLUSION

Many Americans share "a common belief that, today, people must be asked for permission for each and every release of their health information."[367] They are mistaken. At all times in our nation's history, there have been pathways for nonconsensual use of health data. The Institute of Medicine recently recommended moving away from a consent-based model altogether for certain types of health informational research and replacing it with two alternatives: one would rely on certified entities, operating under strict privacy and information security requirements, to manage data uses; the other would rely on "waiver of informed consent by an ethics oversight board."[368] The waiver provisions of current regulations were never designed to serve as the gateway for nonconsensual use of data and they have multiple flaws. Data propertization will not solve these problems.

Psychologists have observed that feelings of ownership "are so basic to the human psyche that communities will create rudimentary property rights even in the absence of formal legal structures."[369] Modern utilitarian property theory has not fully eradicated the popular conception of property "as an extension of the human person."[370] This personhood-based account of property is implicit in the tendency to link property and privacy[371] and may account for the strong urge people feel to consider ownership as a way to address data privacy and access issues. This urge must be resisted. It distracts from the more important questions, "What is an appropriate public use of private data?" and "How shall we make that decision?"

---

366 *See* Jody Freeman, *Extending Public Law Norms Through Privatization*, 116 HARV. L. REV. 1285 (2003) (discussing public law norms that should apply to private decision-making bodies in the context of public-private partnerships).

367 65 *Fed. Reg.* at 82472 (discussing, in the preamble to the 2000 version of the HIPAA Privacy Rule, comments from members of the public who held this belief).

368 IOM, PRIVACY REPORT, *supra* note 4, at 267 (calling for a framework for research use of data that would move away from individual consent requirements in certain circumstances, instead relying on two alternative mechanisms: (1) the use of certified entities that would manage data uses subject to strict privacy and data security requirements, or (2) waiver of informed consent by an ethics oversight board).

369 Bell, *supra* note 50, at 528.

370 Claeys, supra note 49, at 1560.

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