ROBERT H. MCKINNEY SCHOOL OF LAW

INDIANA UNIVERSITY

Indianapolis

Robert H. McKinney School of Law
Legal Studies Research Paper No. 2013 - 04

Nicolas Terry

Protecting Patient Privacy in the Age of Big Data

PROTECTING PATIENT PRIVACY IN THE AGE OF BIG DATA*

By Nicolas P. Terry**

I. INTRODUCTION

The new Administration will determine the future of privacy protection in the United States. At first sight, much the same could have been said of *all* the administrations of the last five or six decades. In each case, the incoming president could have stepped up to the plate and made privacy a legislative or regulatory priority issue. Yet, none did (although a nod is due to the Clinton administration for its HIPAA rules). This time, however, the stakes are different. Failure to act during the next four years will send an almost irrefutable signal to the data collection and aggregation industries that "big data" cannot be stopped or even slowed. As explained below, "big data" refers to a revolution in data collection and processing that dramatically increases the privacy risks imposed on data subjects.

This article takes the position that, beyond its generalized threat to privacy, big data poses an exceptional group of problems for health care, its providers, researchers, and patients. Rightly or wrongly, policymakers have agreed that patient information is deserving of elevated protection compared to other data (so-called health privacy exceptionalism). Yet, at the same time, the last two administrations, one Republican and one Democrat, have promoted the dramatic growth of electronic medical records ("EMR")[1] with the specific goal of increasing the collection of clinical data and its broad sharing. As recently noted by the Institute of Medicine ("IoM"), "the U.S. health care system now is characterized by more to do, more to know, and more to manage than at any time in history."[2] Technology, not surprisingly, is viewed as holding the solution because "[a]dvances have made vast computational power affordable and widely available, while improvements in connectivity have allowed information to be accessible in real time virtually anywhere" affording "the potential to improve health care by increasing the reach of research knowledge, providing access to clinical records when and where needed, and assisting patients and providers in managing chronic diseases."[3]

But, while policymakers are staking health care *progress* on big data, they seem less concerned about existential threats to the privacy of health

---

** Hall Render Professor of Law & Co-Director of the Hall Center for Law and Health, Indiana University Robert H. McKinney School of Law. Email: npterry@iupui.edu.

[1] Nicolas P. Terry, *Meaningful Adoption: What We Know or Think We Know About the Financing, Effectiveness, Quality, and Safety of Electronic Medical Records*, 34 J. LEGAL MED. 7 (2013).

[2] INST. OF MED., BEST CARE AT LOWER COST: THE PATH TO CONTINUOUSLY LEARNING HEALTH CARE IN AMERICA S-4 (2012), *available at* http://books.nap.edu/openbook.php?record_id=13444&page=4

[3] *Id.* at 4-1 to 4-2.

information. The ramifications of big data are manifold. Perhaps two examples will serve to explain the thrust of this article. First, our "medical selves" exist outside of the traditional (and HIPAA/HITECH-regulated[4]) health domain, creating exploitable confusion as our health information moves in and out of protected spaces. Second, big data positions data aggregators and miners to perform an end-run around health care's domain-specific protections by creating medical profiles of individuals in HIPAA-free space. After all, what is the value of HIPAA/HITECH sector-specific protection designed to keep unauthorized data aggregators out of our medical records if big data mining allows the creation of surrogate profiles of our medical selves?

Fortunately, health information technologies ("HIT") and patient privacy share a long history of bipartisan support[5] and the second Obama administration will need to leverage that tradition to protect patients and their sensitive information in the face of growing data aggregation and sophisticated data mining.[6] This battle has to be fought on three fronts. First, while HIPAA/HITECH provide increasingly robust protections against unauthorized uses of health information by a relatively narrow set of traditional health care provider data stewards, it does almost nothing to regulate the *collection* of health data. This is because the HIPAA Privacy Rule is a misnomer. It is not a privacy rule because it only protects against data disclosure not against data collection. It

---

[4] HIPAA/HITECH regulation consists of Privacy and Security Rules made under the Health Insurance Portability and Accountability Act of 1996 (HIPAA) as amended by the Health Information Technology for Economic and Clinical Health Act (HITECH) Act (2009). *See* Health Insurance Portability and Accountability Act (HIPAA) of 1996, Pub. L. No. 104-191, 110 Stat. 1936 (codified as amended in scattered sections of 42 U.S.C.); Health Information Technology for Economic and Clinical Health (HITECH) Act, Pub. L. No. 111-5, 123 Stat. 115, 226 (2009) (codified as amended in scattered sections of 42 U.S.C.). The combined rules are available at U.S. DEP'T OF HEALTH & HUM. SERVS., OFFICE FOR CIVIL RIGHTS, HIPAA ADMINISTRATIVE SIMPLIFICATION (2006), http://www.hhs.gov/ocr/privacy/hipaa/administrative/privacyrule/adminsimpregtext.pdf. However, this will be replaced in early 2013 by an omnibus rule that contains additional regulatory changes made under the authority of the HITECH Act. *See* Modifications to the HIPAA Privacy, Security, Enforcement, and Breach Notification Rules Under the Health Information Technology for Economic and Clinical Health Act and the Genetic Information Nondiscrimination Act, 78 Fed. Reg. 5566 (proposed Jan. 15, 2013) (to be codified at 45 C.F.R. pts. 160 & 164), *available at* http://www.gpo.gov/fdsys/pkg/FR-2013-01-25/pdf/2013-01073.pdf.

[5] One of the more effective groups promoting a somewhat depoliticized approach to promoting HIT and the protection of health information is the Bipartisan Policy Center ("BPC"). *See, e.g.*, Transforming Health Care: The Role of Health IT, BIPARTISAN POL'Y CENTER (Jan. 27, 2012), http://www.bipartisanpolicy.org/news/multimedia/2012/02/transforming-health-care-role-health-it. The BPC is led by Tom Daschle and Bob Bennett, two influential former U.S. senators. *See* Bernie Monegain, *Health IT a Bridge Between Democrat Daschle, Republican Bennett*, GOV'T HEALTH IT (Apr. 27, 2011), http://www.govhealthit.com/news/health-it-bridge-between-democrat-daschle-republican-bennett.

[6] *See generally* Mike Miliard, *Will Health IT Bipartisanship Survive the Elections?*, GOV'T HEALTH IT (Feb. 06, 2012), http://www.govhealthit.com/news/will-health-it-bipartisanship-survive-elections (discussing bipartisan support of health IT).

is therefore more appropriately described as a confidentiality rule.[7] In the world of big data this is like bringing the proverbial knife to a gunfight. As a result it is time that the federal government put real limits on the collection and processing of personal information.

Second, the United States has adopted a sector-based approach to data protection.[8] HIPAA, as amended by HITECH, and the "privacy" and security regulations made thereunder apply only to a narrowly constructed version of the vertical health care market.[9] Such sector-based approaches to regulation are frequently flawed because of poor calibration. This is the case with health information. The health care sector and its stakeholders constitute an area considerably larger than the HIPAA-regulated zone. As a result, some traditional health information circulates in what may be termed a HIPAA-free zone. Further, the very concept of health sector specific regulation is flawed because health related or medically inflected data[10] frequently circulates outside of the traditionally recognized health care sector. In both cases agreed-upon health privacy exceptionalism is jeopardized.

Third, the IoM is correct that there is great value in patient information that could be extracted and used by responsible medical and public health researchers.[11] Responsible public policy suggests that researchers should be able to request that information from patients.[12] Many or most of the existing HIPAA and HITECH security and confidentiality protections will apply here. But neither current policy nor regulation provide the key component: a coherent choice architecture for dealing with appropriate patient decision-making regarding research use of personal or familial health data.

In suggesting legislative amelioration of these three issues, this article does not propose an exhaustive overhaul of HIPAA/HITECH. Rather, it suggests an incremental and additive approach. This includes adopting aspects of two privacy proposals recently published by the White House and the Federal Trade Commission ("FTC"). Part I of this article describes big data and implicated health data pools. Part II discusses some of the meta questions and policy choices applicable to health privacy. Thereafter, Parts III, IV, and V discuss the three core issues requiring resolution (respectively: regulating collection, fixing

---

[7] *See* Nicolas P. Terry & Leslie P. Francis, *Ensuring the Privacy and Confidentiality of Electronic Health Records*, 2007 U. ILL. L. REV. 681, 708 (2007).

[8] For example, the financial sector is regulated by the Gramm-Leach-Bliley Act, Pub. L. No. 106-102, 113 Stat. 1338 (1999).

[9] For example, primarily HIPAA is concerned only with a sub-set of health care providers called "covered entities." 45 C.F.R. § 160.103.

[10] The first person I heard use this phrase was Frank Pasquale. *See* Frank Pasquale, SETON HALL LAW, http://law.shu.edu/Faculty/fulltime_faculty/Frank-Pasquale.cfm (last visited Jan. 30, 2013).

[11] INST. OF MED., *supra* note 2, at S-9, S-13.

[12] *Id.* at S-13 to S-15.

the flaws caused by sector-based regulation, and designing a choice architecture for patient consent to beneficial research).

II. BIG DATA AND HEALTH CARE DATA POOLS

The health care privacy endgame has always been about more than nurses peeking inside a celebrity's HIPAA-protected record.[13] Several years ago Leslie Francis and I cautioned that electronic medical records ("EMRs"):

> [M]ay be searched in problematic ways. Records might be accessible to those who many believe should not have access to them (secondary users). Commercial entities may seek to add medical data to their other data holdings and sell the aggregated data for marketing or surveillance purposes. Groups might be targeted in epidemiological searches.[14]

In the words of a recent McKinsey report, "[l]ike other essential factors of production such as hard assets and human capital, it is increasingly the case that much of modern economic activity, innovation, and growth simply couldn't take place without data."[15]

Big data is beguiling. Like so many of the phenomena we confront in our information society, it promises benefits for almost no cost. If information can move around freely then transactions lose friction and everyone wins. We accept that transactional friction and related inefficiencies are major barriers to health care improving its woeful cost and quality issues. Add in the broadly expressed sentiment that increased application of information technologies is a potential solution.[16] The almost inevitable conclusion is that health care should join the big data revolution. Writing on the *Health Affairs* blog, Robert Kocher and Bryan Roberts describe the upside:

> Big data offers great potential. In other sectors, mining large data sets has led to breakthroughs in productivity, consumer experience, and cost structure. It is also needed to make approaches such as IBM's Watson super computer practical for healthcare, as the quality of machine learning results depends substantially on the amount of data available.

---

[13] *See 27 Suspended for Clooney File Peek*, CNN (Oct. 10, 2007), http://articles.cnn.com/2007-10-10/entertainment/clooney.records_1_hipaa-health-insurance-portability-palisades-medical-center?_s=PM:SHOWBIZ.

[14] Terry & Francis, *supra* note 7, at 683.

[15] JAMES MANYIKA ET AL., MCKINSEY GLOBAL INST., BIG DATA: THE NEXT FRONTIER FOR INNOVATION, COMPETITION, AND PRODUCTIVITY, preface (2011), http://www.mckinsey.com/~/media/McKinsey/dotcom/Insights%20and%20pubs/MGI/Research/Technology%20and%20Innovation/Big%20Data/MGI_big_data_full_report.ashx.

[16] Nicolas P. Terry, *Information Technology's Failure to Disrupt Healthcare*, 13 NEV. L.J. (forthcoming 2013), *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2118653.

> It will not be long until patient level information is combined with large existing data sets [that] . . . will generate far more accurate predictive modeling, personalization of care, assessment of quality and value for many more conditions, and help providers better manage population health and risk-based reimbursement approaches.[17]

But big data is also big business with annual revenues approaching $34 billion.[18] And those who aggregate and mine this data neither view their informational assets as public goods held on trust nor seem particularly interested in protecting the privacy of their data subjects. The truth lies in the opposite because the big data business model is selling information about their data subjects.

A. The Big Data Model

Not surprisingly, "big" data is frequently defined in terms of its size. It even finds definition from what it is not ("datasets whose size is beyond the ability of typical database software tools to capture, manage, and analyze"[19]) and what it might be (vague estimates as to the petabytes and exabytes[20] of information that are being captured).

Big data is closely linked both literally and by its scale to the massive datasets compiled by well know data aggregators such as ChoicePoint or Acxiom.[21] Those datasets often start by aggregating large (but not "big") structured sets created by state, federal, and local governments, law enforcement, and financial institutions amongst others. Acxiom is reported to hold data on five-hundred million consumers with an average of 1500 data points per data subject.[22]

Increasingly and of considerable importance going forward, big data comes from less structured sources including "[w]eb-browsing data trails, social network communications, sensor data and surveillance data."[23] Much of it is

---

[17] Robert Kocher & Bryan Roberts, *Meaningful Use of Health IT Stage 2: The Broader Meaning*, HEALTH AFF. (Mar. 15th, 2012), http://healthaffairs.org/blog/2012/03/15/meaningful-use-of-health-it-stage-2-the-broader-meaning/.

[18] Jordan Robertson, *The Health-Care Industry Turns to Big Data*, BLOOMBERG BUS. WEEK (May 17, 2012), http://www.businessweek.com/printer/articles/26016-the-health-care-industry-turns-to-big-data.

[19] MANYIKA ET AL., *supra* note 15, at 1.

[20] *Megabytes, Gigabytes, Terabytes . . . What Are They?*, WHAT'S A BYTE?, http://www.whatsabyte.com/ (last visited Jan. 30, 2013).

[21] *See, e.g.*, Natasha Singer, *Mapping, and Sharing, the Consumer Genome*, N.Y. TIMES (June 16, 2012), http://www.nytimes.com/2012/06/17/technology/acxiom-the-quiet-giant-of-consumer-database-marketing.html?_r=2&pagewanted=all.

[22] Natasha Singer, *Consumer Data, But Not for Consumers*, N.Y. TIMES, July 22, 2012, at BU3.

[23] Steve Lohr, *Amid the Flood, A Catchphrase Is Born*, N.Y. TIMES, August 12, 2012, at BU3. *See generally* Sean Gallagher, *Big Brother on a Budget: How Internet Surveillance Got So Cheap*, ARS

"exhaust data," or data created unintentionally as a byproduct of social networks,[24] web searches, smartphones,[25] and other online behaviors.[26] Like the traceable signatures of submarines exploited in so many adventure movies,[27] these data trails are ripe for analytic exploitation. Second, and clearly overlapping with the first, big data increasingly will be derived from *The Internet of Things*:[28]

> [T]he predictable pathways of information are changing: the physical world itself is becoming a type of information system. In what's called the Internet of Things, sensors and actuators [(tiny devices)] embedded in physical objects—from roadways to pacemakers—are linked through wired and wireless networks, often using the same Internet Protocol (IP) that connects the Internet. These networks churn out huge volumes of data that flow to computers for analysis. When objects can both sense the environment and communicate, they become tools for understanding complexity and responding to it swiftly. What's revolutionary in all this is that these physical information systems are now beginning to be deployed, and some of them even work largely without human intervention.[29]

According to a McKinsey report "[m]ore than 30 million networked sensor nodes are now present in the transportation, automotive, industrial, utilities, and retail sectors" while "[t]he number of these sensors is increasing at a

---

TECHNICA (Sept. 21, 2012, 9:50 AM), http://arstechnica.com/information-technology/2012/08/big-brother-meets-big-data-the-next-wave-in-net-surveillance-tech/ (discussing the digital trail left when sending e-mails, searching the Web and posting messages, and how companies use the data).

[24] *See* Nicolas P. Terry, *Fear of Facebook: Private Ordering of Social Media Risks Incurred by Healthcare Providers*, 90 NEB. L. REV. 703, 703-51 (2012); *see also* Nicolas P. Terry, *Physicians and Patients Who "Friend" Or "Tweet": Constructing a Legal Framework for Social Networking in a Highly Regulated Domain*, 43 IND. L. REV. 285, 285-341 (2010).

[25] Jennifer M. Urban, Chris Jay Hoofnagle & Su Li, *Mobile Phones and Privacy* (Univ. of Cal., Berkeley Pub. Law Research, Working Paper No. 2103405, 2012), *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2103405.

[26] *See* MANYIKA ET AL., *supra* note 15, at 1; WORLD ECON. FORUM, BIG DATA, BIG IMPACT: NEW POSSIBILITIES FOR INTERNATIONAL DEVELOPMENT 2-3 (2012), *available at* http://www3.weforum.org/docs/WEF_TC_MFS_BigDataBigImpact_Briefing_2012.pdf; *see also* Sarah Perez, *Amazon's Silk Browser Now Tracking User Behavior For New "Trending Now" Section . . . Wait, What?*, TECHCRUNCH (Sept. 7, 2012), http://techcrunch.com/2012/09/07/amazons-silk-browser-now-tracking-user-behavior-for-new-trending-now-section-wait-what/.

[27] *See, e.g.*, THE HUNT FOR RED OCTOBER (Paramount Pictures 1990).

[28] Kevin Ashton, *That 'Internet of Things' Thing*, RFID JOURNAL (July 22, 2009), http://www.rfidjournal.com/article/view/4986. "[T]oday's information technology is so dependent on data originated by people that our computers know more about ideas than things. If we had computers that knew everything there was to know about things—using data they gathered without any help from us—we would be able to track and count everything, and greatly reduce waste, loss and cost." *Id.*

[29] Michael Chui, Markus Löffler & Roger Roberts, *The Internet of Things*, MCKINSEY Q., Mar. 2010, at 70. Excerpt of the article can be found at: http://www.mckinsey.com/insights/mgi/research/technology_and_innovation/the_internet_of_things.

rate of more than 30 percent a year."[30] Big data is creating a *private* surveillance model that will exceed law enforcement tracking of individuals using Internet and cell phone data.[31]

Technically, "big data" refers both to the ability to store and aggregate these giant datasets and the availability of increasingly powerful data mining and analysis techniques.[32] As explained by Steve Lohr, "[b]ig data is a shorthand label that typically means applying the tools of artificial intelligence, like machine learning, to vast new troves of data beyond that captured in standard databases."[33]

Data aggregation and customer profiling are hardly news. The developments that mark out big data are the scale of the data collection and the increasing sophistication of predictive analytics. For example, Facebook sends user data to Datalogix, which matches them to data about customer purchases from stores in an effort to validate Facebook advertising.[34] Similarly, retail operations like Amazon.com and Wal-Mart use sophisticated modeling to recommend purchases based on consumers' prior behavior.[35] This pivot from mere profiling to behavioral tracking is critical. According to a Deloitte study, the analytical focus is to "look[] at that behavior in the context of who is around them and how they interact. It gets below the curated, surface-level information we put out about ourselves online, to help understand what our digital trails and network really say about us."[36] As explained by Siva Vaidhyanathan, market segmentation is the reason for this increased use of data analysis. The big data collectors and aggregators "are devoted to tracking our eccentricities because they understand that the ways we set ourselves apart from the mass are the things about which we are most passionate" and that is because "[o]ur passions, predilections, fancies, and fetishes are what we are likely to spend our surplus

---

[30] MANYIKA ET AL., *supra* note 15, at 2.

[31] *See, e.g.*, *The End of Privacy?*, N.Y. TIMES, July 14, 2012, at SR10; *see also* Andy Greenberg, *U.S. Government Requests For Google Users' Private Data Jump 37% In One Year*, FORBES (June 17, 2012, 11:01 AM), http://www.forbes.com/sites/andygreenberg/2012/06/17/u-s-government-requests-for-google-users-private-data-spike-37-in-one-year/.

[32] MANYIKA ET AL., *supra* note 15, at 2.

[33] Lohr, *supra* note 23.

[34] Rebecca Greenfield, *Facebook Now Knows What You're Buying at Drug Stores*, ATLANTIC WIRE (Sept. 24, 2012), http://www.theatlanticwire.com/technology/2012/09/facebook-tracking-you-drug-store-now-too/57183.

[35] *See generally* Neal Leavitt, *Will Recommendation Technology Boost E-Commerce?*, LEAVITT COMM. (May 2006), *available at* http://www.leavcom.com/ieee_may06.htm (discussing how online retailers have utilized recommendation technology); THOMAS H. DAVENPORT, BABSON EXEC. EDUC., REALIZING THE POTENTIAL OF RETAIL ANALYTICS: PLENTY OF FOOD FOR THOSE WITH THE APPETITE (2009), *available at* http://analytics.typepad.com/files/retailanalytics.pdf (providing a general overview of the use of analytics for retail marketing).

[36] Doug Palmer, Vikram Mahidhar & Dan Elbert, *Making Sense of Social Data*, DELOITTE REV. (2011), at 8, *available at* http://www.deloitte.com/view/en_US/us/Insights/Browse-by-Content-Type/deloitte-review/f7e9ceb3b2741310VgnVCM1000001a56f00aRCRD.htm#.

cash on and thus make us easy targets for precise marketing."[37] Vaidhyanathan's example was of the music aficionado who identified (segmented) his or her "passion" by appreciating a rock band's bluesier catalog. The reality faced here is the exposure of a patient's differential health status and its exploitation for commercial gain.

However, the big data phenomenon goes far beyond the enormity of traditional data pools and super-computer analysis. It also reflects a paradigm shift in data collection. "In the old, data-is-scarce model, companies had to decide what to collect first, and then collect it," but "[w]ith the new, data-is-abundant model, we collect first and ask questions later."[38] Big data also shakes our understanding of some traditional models of data protection because it is "big enough to raise practical rather than merely theoretical concerns about the effectiveness of anonymization."[39]

Health care has two major intersects with big data. First, big data is touted as a way to improve health care,[40] an issue that is explored below.[41] Second, health data is viewed as a major source of big data. As to the latter, big data targets four large health care data pools: (1) drug and device data, (2) clinical data, (3) claims and related financial data, and (4) "patient behavior and sentiment data."[42]

Big data will eventually pull from all four of these pools. However, in the short term, proprietary concerns likely will override any meaningful sharing of drug and device data by manufacturers or claims and related financial data by health care providers. In contrast, clinical data seems to have more of a will to find its way out and be included in big data. For a start, proprietary/commercial interests are likely to bend in the direction of exploitation rather than proprietary curation. Second, there is a powerful "sharing" meme increasingly surrounding clinical data and, whether directly or through function creep, much is destined to leak out.

Although the eventual migration of clinical data into big data is likely inevitable, it is quite a big ask at the moment. First, although privacy advocates view U.S. Department of Health & Human Services' ("HHS") meaningful use

---

[37] SIVA VAIDHYANATHAN, THE GOOGLIZATION OF EVERYTHING: (AND WHY WE SHOULD WORRY) 112-13 (2011); *see also* Natasha Singer, *Shoppers, Meet Your Scorekeeper*, N.Y. TIMES, Aug. 19, 2012, at BU1 (discussing how e-scores may be used to target potential customers presumed to have buying power).

[38] Alistair, *Big Data Is Our Generation's Civil Rights Issue, and We Don't Know It*, SOLVE FOR INTERESTING (July 31, 2012, 12:40 PM), http://solveforinteresting.com/big-data-is-our-generations-civil-rights-issue-and-we-dont-know-it/.

[39] Howard Wen, *Big Ethics For Big Data*, STRATA (June 11, 2012), http://strata.oreilly.com/2012/06/ethics-big-data-business-decisions.html (quoting Douglas Patterson).

[40] *See* Terry, *supra* note 16.

[41] *See infra* Part VI.

[42] MANYIKA ET AL., *supra* note 15, at 42.

subsidy program for EMRs as the final nail in the coffin of health privacy,[43] the reality is rather different. Other than in a few vertically integrated health systems, clinical data remains quite fragmented[44] while the current generation of EMRs is woefully limited. This is particularly the case with regard to the execution of tasks such as data sharing with patients or interoperability with the data systems of other providers.[45] Second, EMR "leakage" is a health data risk that is actually covered by regulation. In 2009, HITECH introduced regulatory authority relating to accounting provisions for EMR data[46] and limitations on the commercial exploitation of EMR data.[47]

Ironically, and likely only in the near-term, patients currently are the more likely source of the EMR clinical data that migrates into big data pools. As calls for patient engagement increase, services such as lightly regulated Personal Health Records ("PHR")[48] and programs such as the "Blue Button"[49] initiative will proliferate. Patients will be encouraged (often by privacy advocates who view this activity as autonomy-satisfying "control") to download their records from HIPAA-protected EMRs. At that point, the patient-curated copy of the data loses its HIPAA protection and may be subject to exposure like other medically inflected data.

B. Medically Inflected Data and Health Information Surrogates

In the short-term, patient behavior and sentiment data, the final of the four health care data pools described above, will be of the greatest importance. First, in the current business-legal-technical environment, it is the easiest to acquire. Second, when medically inflected data is given the big data treatment the result increasingly will be a valuable surrogate or substitute version for

---

[43] *See* Jerome Groopman & Pamela Hartzband, *Obama's $80 Billion Exaggeration*, WALL ST. J., Mar. 12, 2009, at A15. "Some have speculated that the patient data collected by the Obama administration in national electronic health records will be mined for research purposes to assess the cost effectiveness of different treatments. This analysis will then be used to dictate which drugs and devices doctors can provide to their patients in federally funded programs like Medicare." *Id*.

[44] Michelle McNickle, *5 Basics of Big Data*, HEALTHCARE IT NEWS (June 13, 2012), http://www.healthcareitnews.com/news/5-basics-big-data.

[45] *See* Terry, *supra* note 1, at 25-26, 39.

[46] 42 U.S.C.A. § 17935(c) (West 2010) (corresponds to the HITECH Act, Pub. L. No. 111-5, § 13405(c), 123 Stat. 115, 264-67).

[47] *Id*. § 17935(d) (HITECH Act, § 13405(d), 123 Stat. 115, 266-68).

[48] *See generally* Nicolas P. Terry, *Personal Health Records: Directing More Costs and Risks to Consumers?*, 1 DREXEL L. REV. 216 (2009) (discussing the growth and use of personalized health records).

[49] *See generally* Lygeia Ricciardi & Doug Fridsma, *Call for Participation in the Automate Blue Button Initiative: Enhancing Consumer Access to Health Information*, HEALTH IT BUZZ (Aug. 9, 2012, 6:15 PM), http://www.healthit.gov/buzz-blog/electronic-health-and-medical-records/blue-button-initiative-enhancing-consumer-access-health-information/ (discussing the necessity and demand for individuals to make use of the web tool "Blue Button").

traditional patient health information. Crucially, this surrogate or shadow information will be subject to only the lightest form of data protection (e.g., compliance with published privacy policies) and certainly will be exempt from the exceptional protection provided traditional patient health information by HIPAA/HITECH.

One of the first descriptions of this phenomenon was by Charles Duhigg. He explained how Target Corp. used predictive analytics to identify potential customers in their second trimester of pregnancy so that they could send them targeted advertising.[50] This was achieved primarily by using analytics to discover changes in the patterns of product purchases (e.g., changing to unscented lotion).[51] Writing in *The Atlantic*, William Pewen observed:

> Rapidly expanding and interconnected data thus raise the risk of adverse impacts arising from discrimination, posing the challenge of how to realize the benefits of technology while minimizing harm—and applications involving personal health data present some of the greatest hazards. While considerable risk reduction can be achieved by simply securing health data in accordance with the ancient Hippocratic Oath, the creation of health data surrogates threatens to undermine that.[52]

Such medically inflected data (or patient behavior and sentiment data) increasingly will be sourced from the exhaust data of online transactions. For example, an online bookstore site may have recorded the rental of a book about breast cancer survival rates, a supermarket loyalty card's database may have recorded the purchase of a pregnancy-testing kit, a social media user may have "liked" the Parkinson's Disease Foundation page,[53] or a search engine may have recorded a search for PatientsLikeMe.[54] Additional data will be pulled from the health care specific Internet of Things—such as data from medical devices and the radio-frequency identification chips embedded in drug packaging.[55]

It is unknown exactly how much health information or medically inflected information data aggregators include in their data sets. However,

---

[50] Charles Duhigg, *How Companies Learn Your Secrets*, NY TIMES, Feb. 16, 2012, *available at* http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html?pagewanted=all.
[51] *Id.*
[52] William F. Pewen*, Protecting Our Civil Rights in the Era of Digital Health*, THE ATLANTIC (Aug. 2, 2012, 11:09 AM), http://www.theatlantic.com/health/archive/2012/08/protecting-our-civil-rights-in-the-era-of-digital-health/260343/2/.
[53] *Parkinson's Disease Foundation*, FACEBOOK, http://www.facebook.com/parkinsonsdisease foundation (last visited Jan. 30, 2013).
[54] PATIENTSLIKEME, http://www.patientslikeme.com (last visited Jan. 30, 2013).
[55] *See, e.g.*, Amy Dockser Marcus & Christopher Weaver, *Heart Gadgets Test Privacy-Law Limits*, WALL ST. J. (Nov. 28, 2012, 10:31 PM), http://online.wsj.com/article/SB10001424052970203937 00457807882087474407 6.html (discussing privacy issues regarding data transmitted from an implanted defibrillator back to the device's manufacturer).

Acxiom's own "Consumer Data Products Catalog" lists a number of health or health-related data categories for sale.[56]

C. Legal Consideration of Big Data

Given its scale and the implications of its practices, the big data industry has managed to keep a relatively low legal profile. Overall, big data is lightly regulated. Only occasionally will an aggregator or broker fall afoul of some specific regulatory requirement, for example by failing to comply with the Fair Credit Reporting Act.[57] Additionally, we have seen the occasional Federal Trade Commission ("FTC") case brought against data aggregators who have failed to prevent unauthorized access to sensitive consumer information stored in their databases[58] or data collectors who continued to collect information about consumers who had opted out of such.[59]

Notwithstanding this legal low profile, two recent cases demonstrate the increasing footprint of big health care data. In *Sorrell v. IMS Health, Inc*., the U.S. Supreme Court struck down a Vermont statute that restricted the sale and use of pharmacy records that documented the prescribing practices of physicians.[60] The Court held that the statute imposed content-based restriction on protected speech and further that the state had failed to justify the restriction under the applicable heightened scrutiny standard.[61] Authoring the dissent Justice Breyer seemed less concerned, arguing that the challenged Vermont statute had only one relatively minor effect: "depriv[ing] pharmaceutical and data-mining companies of data, collected pursuant to the government's regulatory mandate, that could help pharmaceutical companies create better sales

---

[56] *Consumer Data Products Catalog*, ACXIOM, http://www.acxiom.com/site-assets/catalog-consumer-data-products/ (last visited Jan. 30, 2013).

[57] *E.g.*, Spokeo settled FTC allegations that its compilations of social media and other data about individuals constituted "consumer reports" and that it breached the Fair Credit Reporting Act by not adequately policing use and accuracy of data. *See Spokeo to Pay $800,000 to Settle FTC Charges Company Allegedly Marketed Information to Employers and Recruiters in Violation of FCRA*, FED. TRADE COMM'N. (June 12, 2012), http://www.ftc.gov/opa/2012/06/spokeo.shtm.

[58] *See, e.g.*, *In re* Reed Elsevier, Inc. & Seisint, Inc., No. 052-3094, 2008 WL 3150420 (F.T.C. 2008). Case documents can also be found at: *In the Matter of Reed Elsevier Inc. and Seisint, Inc.*, FED. TRADE COMM'N, http://www.ftc.gov/os/caselist/0523094/index.shtm (last visited Jan. 30, 2013).

[59] *See, e.g.*, *In the Matter of Chitika, Inc.*, FTC File No. 1023087, FED. TRADE COMM'N, *available at* http://www.ftc.gov/os/caselist/1023087/index.shtm (last visited Jan. 30, 2013). *See generally* Jay Cline, *7 Reasons the FTC Could Audit Your Privacy Program*, COMPUTERWORLD (Aug. 21, 2012, 3:08 PM), http://www.computerworld.com/s/article/9230458/7_reasons_the_FTC_could_audit_your_privacy_program?taxonomyId=84 (discussing the practices that might trigger an FTC investigation of a company).

[60] Sorrell v. IMS Health, Inc., 131 S.Ct. 2653, 2656-58 (2011).

[61] *Id.*

messages."[62] For the data miners, however, even this was too great a restriction on their big data business that the majority opinion described as follows:

> Pharmacies, as a matter of business routine and federal law, receive prescriber-identifying information when processing prescriptions. . . . Many pharmacies sell this information to "data miners," firms that analyze prescriber-identifying information and produce reports on prescriber behavior. Data miners lease these reports to pharmaceutical manufacturers subject to nondisclosure agreements. Detailers, who represent the manufacturers, then use the reports to refine their marketing tactics and increase sales.[63]

This data mining, therefore, is an example of widespread data aggregation and mining involving information that had its origins in information about patients. The privacy implications of the case generally have been downplayed because the patient data apparently was de-identified and the Vermont prohibition only targeted disclosure for marketing purposes. However, Ashutosh Bhagwat has been less sanguine about the opinion's broader privacy implications, arguing that the opinion appears to broadly protect "factual speech."[64] Bhagwat speculates that such protection potentially could hinder privacy regulation of everything from personal health information to the myriad sources of big data.[65]

State of Minnesota v. Accretive Health, Inc., is a very different case, but not only because it ended in a settlement.[66] The case is not particularly important from a traditional privacy perspective, even though the Minnesota Attorney General's investigation was triggered by the alleged theft of an Accretive laptop containing data from about 23,000 patients.[67] Rather, *Accretive* illustrates some of the possible implications of the big data cocktail, mixing clinical with non-clinical data. According to the Attorney General's complaint, Accretive was a

---

[62] *Id*. at 2673.
[63] *Id*. at 2660.
[64] Ashutosh Bhagwat, *Sorrell v. IMS Health: Details, Detailing, and the Death of Privacy*, 36 VT. L. REV. 855, 862-68 (2012).
[65] *Id*. at 868.
[66] Minnesota v. Accretive Health, Inc., No. 12-145 (D. Minn. Aug. 7, 2012); *see also* Press Release, Office of Attorney General Lori Swanson, Attorney General Swanson Says Accretive Will Cease Operations in the State of Minnesota Under Settlement of Federal Lawsuit (July 31, 2012), *available at* http://www.ag.state.mn.us/Consumer/PressRelease/07312012AccretiveCease Operations.asp.
[67] This was also the first example of a state Attorney General leveraging the amended HIPAA enforcement powers introduced in 2009. HITECH § 13410(e) states that "in any case in which the attorney general of a State has reason to believe that an interest of one or more of the residents of that State has been or is threatened or adversely affected by any person who violates a provision of this part, the attorney general of the State . . . may bring a civil action on behalf of such residents of the State." 42 U.S.C.A. § 1320d-5 (West 2010).

licensed debt-collector and, pursuant to contracts, had control of some Minnesota hospitals' "revenue cycles."[68] The apparent upshot was that Accretive had managerial responsibility over hospital employees in charge of everything from scheduling to billing and collection, "'infused' its own employees into the staff of the hospitals . . . [and] engage[d] in 'data mining' and 'consumer behavior modeling' on patients."[69] It soon became apparent that the Attorney General's real concerns were "aggressive collection practices in hospital emergency rooms . . . ."[70] The settlement of the case involved the payment of $2.5 million, Accretive's cessation of operations in the state, and the return of all data to the hospitals.[71] The lesson of the case however, is the way corporate America increasingly looks to collect data from multiple sources and then leverages it. The *Accretive* case involved relatively "small" data but the admixture model is illustrative of the path being followed by big data aggregators and miners as they increasingly combine data pools for commercial advantage.

As we will see, however, neither sector-based nor general privacy protections are prepared for this onslaught. Equally, *Sorrell*, together with other recent Supreme Court decisions pioneering arguments that favor businesses facing regulation of their information activities,[72] suggest that the regulation of big data will only become more challenging.

III. REFINING PRIVACY POLICY CHOICES

U.S. law generally has been more willing to make broad pronouncements about limitations on government intrusions into the lives of citizens than to protect them from commercial interests coveting their personal information. In a recent Fourth Amendment case, *United States v. Jones*, the surveillance use of a GPS device was successfully challenged.[73] Such surveillance is somewhat analogous to the "breadcrumbs" tracking favored by big data. In her concurrence, Justice Sotomayor noted the interests at stake when technology "generates a precise, comprehensive record of a person's public movements that

---

[68] Complaint at 2, Minnesota v. Accretive Health, Inc., No. 12-145 (D. Minn. Aug. 7, 2012), *available at* http://www.ag.state.mn.us/PDF/Consumer/SecondAmendedSupplementaComplaint.pdf.

[69] *Id.*; *see also* Press Release, Office of Attorney General, *supra* note 66.

[70] Press Release, Office of Attorney General, *supra* note 66.

[71] *Id.*

[72] *See, e.g.*, Citizens United v. Fed. Election Comm'n, 558 U.S. 310 (2010) (declaring unconstitutional the government's suppression of political speech based on the corporate identity of the speaker); R.J. Reynolds Tobacco Co. v. Food & Drug Admin., 845 F. Supp. 2d 266, 274 (D.C. Cir. 2012) (striking down legislation requiring cigarette company to display graphic images on packages of cigarettes).

[73] United States v. Jones, 132 S. Ct. 945, 949 (2012).

reflects a wealth of detail about her familial, political, professional, religious, and sexual associations."[74] And it is a record capable of causing great harm:

> Awareness that the Government may be watching chills associational and expressive freedoms. And the Government's unrestrained power to assemble data that reveal private aspects of identity is susceptible to abuse. The net result is that GPS monitoring—by making available at a relatively low cost such a substantial quantum of intimate information about any person whom the Government, in its unfettered discretion, chooses to track—may "alter the relationship between citizen and government in a way that is inimical to democratic society."[75]

Private tracking performed for commercial purposes is no less harmful. David Rosen refers to the sourcing of big data from government agencies and entities, telecommunications companies, the Internet, and private data aggregators as part of the "corporate-state surveillance complex."[76] Yet, even though the big data industry is creating a private surveillance model that is no less Orwellian than the governmental intrusions discussed in *Jones*, private law privacy protections have been more modest and increasingly sector-based.[77]

A. Core Privacy Models

Data protection regulation is not a monolith built around a single protective principle or rule. Rather, it is a multi-layered construct that can be built from a menu of choices, each of which, if legislated, would in some way constrict the movement of "private" data. The basic menu of choices for policymakers contains:

1. Processes or rules designed to reduce the value of data, thus making collection unlikely (or at least harmless). Such can be achieved, for example, by permitting collection of a type of data only if anonymized or by requiring de-identification prior to processing.
2. Rules that place formal limitations on data collection. Examples would include outright prohibitions on the collection of a particular type of data by particular classes of persons such as by the Genetic Information Nondiscrimination Act ("GINA")[78] or more amorphous prohibitions on

---

[74] *Id.* at 955.
[75] *Id.* at 956 (quoting United States v. Cuevas-Perez, 640 F.3d 272, 285 (7th Cir. 2011)).
[76] David Rosen, *Four Ways Your Privacy is Being Invaded*, SALON (Sept. 11, 2012, 8:44 AM), http://www.salon.com/2012/09/11/four_ways_your_privacy_is_being_invaded/.
[77] *See generally* ALAN F. WESTIN, PRIVACY AND FREEDOM (1970); *see also* Nicolas P. Terry, *What's Wrong with Health Privacy?*, 5 J. HEALTH & BIOMEDICAL L. 1, 3-5 (2009) (discussing limitations of Restatement (Second) of Torts § 652A(2) in terms of the "right of privacy").
[78] Genetic Information Nondiscrimination Act of 2008, Pub. L. No. 110-233, 122 Stat. 881.

the collection of data other than necessary for the transaction in question.

3. Formal limits on the processing of collected data. For example, the law might impose a restriction on using data for purposes other than those it was collected for or the sale of certain types of data (known as a market inalienability rule).[79]
4. Security requirements specifying physical and technological barriers protecting collected data. Such barriers may be aimed at outsiders who would break into the data or insiders who would otherwise breach rules regarding access to or distribution of data.
5. Regulating or otherwise restricting the disclosure or distribution of collected information. For example, a confidentiality rule that requires certain defined classes of data custodians (such as "covered entities" under HIPAA/HITECH) to restrict access to collected data to specific persons or only to cases where the data subject has authorized the disclosure.
6. Breach notification. In cases where data has been put in jeopardy, such that there is a likelihood of unauthorized disclosure, the data custodian is under a duty to notify the data subject and usually some regulatory body.

Several points can be made about this list of legal data protection models. It generally represents a hierarchy of protection. For example, the strongest data protections, towards the top of the list, feature throttling the collection of data. In contrast, the final choice (a *breach notification* rule) is triggered only when data protection has failed and requires an admission that the proverbial horse has been stolen from the stable.[80]

Next, there are several different protective models represented here. The first two choices are true "privacy" rules aimed at regulating the *collection* of data. The third is somewhat of a hybrid in that it imposes processing limits. Those limits will generally make data collection less attractive and so it tends to throttle collection. The fourth menu item is a *security* rule and the sixth, as already discussed, is somewhat of an admission that security has failed. The fifth rule is often described as a privacy rule. However, as noted above, that is an error because privacy models control data *collection.* Yet, this fifth model only controls data *disclosure*. Therefore, it is more accurately described as a confidentiality rule.

---

[79] *See* Margaret Jane Radin, *Market-Inalienability*, 100 HARV. L. REV. 1849, 1853 (1987).

[80] Final rule on Breach Notification for Unsecured Protected Health Information under the HITECH Act, 45 C.F.R. §§ 164.402-410 (included in Omnibus Rule, 78 Fed. Reg. 5,566 (Jan. 25, 2013).

Finally, existing sector-based privacy protection of health information tends to pick from the later, less restrictive models in the table. HIPAA does have a *security* rule[81] and, as discussed herein, *breach notification* rules.[82] But, notwithstanding the fact that it labels its primary data protection rule as a *privacy* rule,[83] HIPAA/HITECH primarily relies on a *confidentiality* model.

HIPAA's confidentiality model has been much criticized.[84] Further, much of its HITECH amelioration involved the adoption of breach notification, a relatively weak privacy model suggesting that the regulators are moving in the wrong direction. Nevertheless, as it exists today incorporating its HITECH tweaks and reflecting more serious enforcement, the HIPAA privacy rule is quite a powerful confidentiality rule. The problem is that in the face of big data, even a good rule policing disclosures likely will prove insufficient. Indeed, a key position taken in this article is that, faced with big data, U.S. policymakers must reverse their journey *down* the menu and adopt the more restrictive models that appear at the top, true privacy controls on data collection.

In 2012, in the space of just a few weeks, the publication of two major reports on privacy suggested that at least some key policymakers are in agreement. In February, the White House released its report *Consumer Data Privacy in a Networked World*,[85] followed in March by the FTC's *Protecting Consumer Privacy in an Era of Rapid Change*.[86] The White House report is the more ambitious of the two. It includes a Consumer Privacy Bill of Rights built around "Fair Information Practice Principles" ("FIPPs").[87] The FTC proposals, while more limited, are a broader mixture of proposals (and current initiatives) and may gain more traction. Those proposals are components in the FTC's "Privacy Framework" containing best practices for data stewards and processors and to make specific legislative proposals.[88] Both sets of proposals are discussed in the sections that follow.

---

[81] 45 C.F.R. pts. 160, 164 (A), (C); *see also Health Information Privacy*, U.S. DEP'T OF HEALTH & HUM. SERVS., http://www.hhs.gov/ocr/privacy/hipaa/administrative/securityrule/index.html (last visited Jan. 30, 2013).

[82] *See* Breach Notification for Unsecured Protected Health Information, 74 Fed. Reg. 42740-01 (Aug. 24, 2009) (codified at 45 C.F.R. pts. 160 & 164).

[83] *Health Information Privacy*, U.S. DEP'T OF HEALTH & HUM. SERVS., http://www.hhs.gov/ocr/privacy/hipaa/understanding/summary/index.html (last visited Jan. 30, 2013).

[84] *See, e.g.*, Terry, *supra* note 77, at 1-32.

[85] THE WHITE HOUSE, CONSUMER DATA PRIVACY IN A NETWORKED WORLD: A FRAMEWORK FOR PROTECTING PRIVACY AND PROMOTING INNOVATION IN THE GLOBAL DIGITAL ECONOMY (2012), *available at* http://www.whitehouse.gov/sites/default/files/privacy-final.pdf [hereinafter CONSUMER DATA PRIVACY].

[86] FED. TRADE COMM'N, PROTECTING CONSUMER PRIVACY IN AN ERA OF RAPID CHANGE: RECOMMENDATIONS FOR BUSINESSES AND POLICYMAKERS (2012), *available at* http://www.ftc.gov/os/2012/03/120326privacyreport.pdf.

[87] CONSUMER DATA PRIVACY, *supra* note 85, at 1-2, 9-22.

[88] FED. TRADE COMM'N, *supra* note 86, at vii.

B. Health Privacy Exceptionalism

Policymakers, patients, and physicians (if not always health care institutions) seem to accept the principle of health privacy exceptionalism—that health information deserves a higher level of privacy protection than most other types of data. The cynic might observe that this exceptional protection was a product of chance, that the health care industry needed HIPAA's electronic data exchange badly enough to agree to privacy regulation. Notwithstanding, because this article argues for strengthening this model by adding additional legal protections, it is important to reiterate some of the reasoning behind health privacy exceptionalism.

Essentially, this exceptionalism is derived from the core value of autonomy, aspects of which mark us out as individual and free. Autonomy is implicated because of the intimate nature of the data being collected and the manner much of it is collected. Thus, Beauchamp and Childress assert this autonomy basis for health privacy as follows: "We owe respect in the sense of deference to persons' autonomous wishes not to be observed, touched, intruded on, and the like. The right to authorize or decline access is basic."[89]

Thereafter, the relationship between health privacy and confidentiality is best understood in terms of chronology. In health care, we exercise our privacy rights by surrendering information to our physician to better our care, and our physician takes that information on trust and promises confidentiality. Or, in the words of Beauchamp and Childress, "[w]e surrender some of our privacy when we grant others access to our personal information or our bodies, but we usually retain some control over information generated about us in diagnostic and therapeutic contexts and in research. . . . When others gain access to protected information without authorization, they infringe our right to confidentiality, our right to privacy, or both."[90] This exceptional model is reinforced by contemporary phenomena such as the increase in medical identity theft[91] and the fact that those stolen medical identities are worth more than regular ones.[92]

---

[89] TOM L. BEAUCHAMP & JAMES F. CHILDRESS, PRINCIPLES OF BIOMEDICAL ETHICS 298 (6th ed. 2009).
[90] *Id.* at 302.
[91] *See Medical Identity Theft and Data Mismanagement*, PRICEWATERHOUSE COOPERS, http://www.pwc.com/us/en/healthcare/publications/medical-identity-theft-and-data-mismanagement.jhtml (last visited Jan. 30, 2013).
[92] Michael Bruemmer, *Vigilance is Still the Best Medicine for Avoiding Medical Identity Theft*, EXPERIAN (June 26, 2012), http://www.experian.com/blogs/data-breach/2012/06/26/vigilance-is-still-the-best-medicine-for-avoiding-medical-identity-theft/; *see also* Jim Avila & Serena Marshall, *Your Medical Records May Not Be Private: ABC News Investigation*, ABC NEWS (Sept. 13, 2012), http://abcnews.go.com/Health/medical-records-private-abc-news-investigation/story?id=17228986#.UGR-IbTA1Y4.

For all the criticisms leveled at it,[93] HIPAA *has* promoted a culture of privacy among health care providers. Like most purely regulatory structures, it has done so less by creating rights in data subjects (HIPAA does the opposite, seeming to establish providers and the state as having primary claims on personal health information) and by creating compliance-driven, rather than conceptual, buy-in by data stewards. Proposals have been made to amend HIPAA (successfully in the case of HITECH) and no doubt there will be more tweaking in the future. It also seems to be the case that HHS (or at least HHS under the present administration) is now committing to more of a balance between compliance and enforcement with the latter being ramped up.[94]

Even with the amendments introduced in HITECH, requiring for example authorizations for the sale of EMR data and tightening up the minimum necessary rule, the regulatory provisions still only apply downstream of collection and within the familiar HIPAA framework of broad exceptions and function creep. In such an environment, privacy frequently is endangered by instrumentalism. The recent McKinsey analysis of big data may have been channeling emerging commercial interest when it noted that "[s]takeholders . . . do not often share existing datasets, because of legal constraints such as privacy laws, a lack of incentives, or incompatible IT systems or formats. . . . *As using big data becomes more important to the industry*, policy makers may have to reevaluate these laws or intervene to ensure that access to data is available in a safe and secure way that also enables health care outcomes to be optimized."[95]

As a result, HIPAA's health data exceptionalism model based on confidentiality needs to be bolstered with other models of data protection. Specifically, privacy models limiting the collection of health information should be introduced channeling some of the recent proposals from the White House and the FTC.

---

[93] *See, e.g.*, Terry & Francis, *supra* note 7, at 683-84 (discussing the advantage of electronic record keeping and doctors' skepticism of HIPAA); Terry, *supra* note 77, at 32 (discussing deficiencies with medical privacy); Terry, *supra* note 48, at 216 (discussing personal health records).

[94] *See, e.g.*, News Release, U.S. Dep't of Health & Human Services, HHS Imposes a $4.3 Million Civil Money Penalty for Violations of the HIPAA Privacy Rule (Feb. 22, 2011), *available at* http://www.hhs.gov/news/press/2011pres/02/20110222a.html (discussing the first civil money penalty issued by HHS for a HIPAA violation); News Release, U.S. Dep't of Health & Human Services, HHS Settles HIPAA Case With BCBST for $1.5 Million (Mar. 13, 2012), *available at* http://www.hhs.gov/news/press/2012pres/03/20120313a.html (discussing first enforcement of the Breach Notification Rule); News Release, U.S. Dep't of Health & Human Services, Massachusetts Provider Settles HIPAA Case for $1.5 Million (Sept. 17, 2012), *available at* http://www.hhs.gov/news/press/2012pres/09/20120917a.html (discussing settlement for potential violation of HIPAA's Security Rule).

[95] MANYIKA ET AL., *supra* note 15, at 52 (emphasis added).

IV. COLLECTION-BASED PRIVACY REGULATION

Writing about the passage of the Genetic Information Nondiscrimination Act,[96] William Pewen noted "two key principles, in confronting . . . technology-based discrimination;"[97] those arguments are broadly applicable to the protection of health information:

> The first is that one cannot rely solely on prohibiting harmful acts. . . . [W]henever Congress moves to prohibit harmful acts, that can trigger a seemingly endless game of "Whack a Mole" as one attempts to identify every conceivable loophole that might be used to undermine legislative intent.
>
> One may thus conclude that reliance on explicit prohibitions alone can be insufficient, and that leads to a second principle that is critical to address discrimination involving health information: that access to health data must be restricted to those with a clear justification for its use. In this case, opening genetic data to employers and others could create a race between creative discrimination and congressional response, and given the current efficacy of Congress, the winner in such a contest appears obvious.[98]

Pewen's call to restrict access to health information is consistent with the sentiment expressed by Beauchamp and Childress that "[i]mproved security . . . will likely address only a few of the interests traditionally protected by rules of confidentiality."[99] Such concerns about the level of data collection are compounded by Mark Rothstein's reminder that "once health information is entered into an EHR, it never goes away, regardless of whether the information continues to be medically relevant or its degree of sensitivity."[100]

Policing data storage and processing with a confidentiality rule is always going to suffer at the hands of "function creep,"[101] instrumental justifications, and error-induced data leaks. As a result, the only effective way to deal with the big data threat is to reduce the amount of data being collected. That requires a true collection-regulating privacy model not a lesser confidentiality model labeled as one, the case with HIPAA/HITECH. The White House's Consumer Privacy Bill of Rights lists six FIPPs,

---

[96] Genetic Information Nondiscrimination Act of 2008, Pub. L. No. 110-233, 122 Stat 881.
[97] Pewen, *supra* note 52.
[98] *Id.*
[99] BEAUCHAMP & CHILDRESS, *supra* note 89, at 304.
[100] Mark A. Rothstein, *Access to Sensitive Information in Segmented Electronic Health Records*, 40 J.L. MED. & ETHICS 394, 394 (2012).
[101] *See generally* Dan Goodman, *Function Creep: Surveillance in London*, THE SAMOVAR (July 17, 2007, 1:42 PM), http://thesamovar.wordpress.com/2007/07/17/function-creep-surveillance-in-london/ (explaining that function creep is "[w]hen a new technology is introduced to do one thing (one function), and is later used for an entirely different thing . . . .").

> − Individual Control: Consumers have a right to exercise control over what personal data companies collect from them and how they use it.
> − Transparency: Consumers have a right to easily understandable and accessible information about privacy and security practices.
> − Respect for Context: Consumers have a right to expect that companies will collect, use, and disclose personal data in ways that are consistent with the context in which consumers provide the data.
> − Security: Consumers have a right to secure and responsible handling of personal data.
> − Access and Accuracy: Consumers have a right to access and correct personal data in usable formats, in a manner that is appropriate to the sensitivity of the data and the risk of adverse consequences to consumers if the data is inaccurate.
> − Focused Collection: Consumers have a right to reasonable limits on the personal data that companies collect and retain.
> − Accountability: Consumers have a right to have personal data handled by companies with appropriate measures in place to assure they adhere to the Consumer Privacy Bill of Rights.[102]

The legal model contemplated for the Bill of Rights is federal legislation producing "legally enforceable codes of conduct" with specific enforcement authority vested in the FTC.[103] Although the White House report called for immediate consultation with stakeholder and implementation of this framework "without delay"[104] and the Department of Commerce has begun to move forward on stakeholder input,[105] it is likely that major work on the proposal has been shelved until the second term.[106]

The FTC Framework contains three "baseline principle[s]," whereby commercial entities should: (1) practice "[p]rivacy by [d]esign" and "promote consumer privacy throughout their organizations and at every stage of the development of their products and services," (2) simplify consumer choice, and (3) "increase the transparency of their data practices."[107]

Such legislation likely would grant the FTC regulatory authority over security policies and practices, breach notification, access and correction for data subjects, and prohibitions against pretextual or other dubious data acquisition tactics.[108] As the agency notes, "the principles . . . are intended to work together

---

[102] CONSUMER DATA PRIVACY, *supra* note 85, at 1. Appendix A of the report contains the full text of the Bill of Rights. *Id.* at 47-48.
[103] *Id.* at 2.
[104] *Id.* at 3.
[105] Multistakeholder Process to Develop Consumer Data Privacy Codes of Conduct, 77 Fed. Reg. 13098-01 (Mar. 5, 2012).
[106] *See infra* Part VII.
[107] FED. TRADE COMM'N, *supra* note 86, at i, vii-viii, 22.
[108] *See, e.g.*, Data Security and Breach Notification Act of 2011, S. 1207, 112th Cong. § 2-3 (2011).

to shift the burden for protecting privacy away from consumers and to encourage companies to make strong privacy protections the default."[109] The FTC's approach is to make commercial entities internalize the data privacy costs of their businesses and technologies by demanding "privacy by design," a process-based approach to lessening the privacy costs of consumers.[110]

Adopting just three of the White House's proposed FIPPs would dramatically improve the protection of patient privacy in the age of big data. Together, "respect for context" and "focused collection" would slow the collection and retention of peripheral or exhaust data potentially impacting health privacy.[111] "Individual control" would "provide consumers appropriate control over the personal data that consumers share with others and over how companies collect, use, or disclose personal data."[112]

The FTC proposals can be criticized for being more powerful in their message, "Privacy by Design," than in their specific proposals for collection-centric regulation. However, "Privacy by Design," is usefully consistent with HHS's newly adopted "Culture of Privacy."[113] The sloganized imperative to increase patient privacy should be applied along with the "substantive principles" that lie behind it, specifically that "Companies should incorporate substantive privacy protections into their practices, such as data security, reasonable collection limits, sound retention practices, and data accuracy."[114]

V. REFORMING SECTOR-BASED REGULATION

HIPAA is formally sector-based, meaning it applies only to health care "covered entities."[115] However, HIPAA-protected information can migrate to unprotected or lightly protected zones.[116] Worse and as already noted, supercomputer class analytics run against medically inflected data may create medical records surrogates in unregulated space.[117]

As a result, future regulation must throw off its industry-specific limitations. Although, as noted above, its trajectory within available privacy models may have been flawed,[118] HITECH did succeed in introducing some limited protection of health information held by non-health care industry custodians. HITECH's breach notification rules not only apply to HIPAA's

---

[109] FED. TRADE COMM'N, *supra* note 86, at 24.
[110] *Id.* at 22.
[111] CONSUMER DATA PRIVACY, *supra* note 85, at 15-17, 21.
[112] *Id*. at 47.
[113] PatientPrivacyRights, *Farzad Mostashari - 2nd International Summit on the Future of Health Privacy*, YOUTUBE (July 17, 2012), http://www.youtube.com/watch?v=foWCyP9zAac.
[114] FED. TRADE COMM'N, *supra* note 86, at 23.
[115] 45 C.F.R. § 160.103 (2012).
[116] *See supra* text accompanying notes 48-49.
[117] *See supra* text accompanying note 50.
[118] *See* Terry, *supra* note 77.

covered entities and business associates,[119] but also to PHR providers who are not covered entities.[120] Similarly the *Personal Health Record Model Privacy Notice* introduced by ONC in late 2011[121] provides "a standardized template that a web-based PHR company can use to succinctly inform consumers about its privacy and security policies"[122] without industry group specific limitation.

Unfortunately, the White House and FTC both seem to be pedaling in the opposite direction, apparently intent on maintaining rigid and disparate sector-based regulation. The White House report, while calling for Congress to enact legislation that includes the Consumer Privacy Bill of Rights, limits that request "to commercial sectors that are not subject to existing Federal data privacy laws."[123] The report is somewhat ambiguous as to whether it recommends extending protection of, say, health information curated by non-health care entities, as it notes that "[t]o avoid creating duplicative regulatory burdens, the Administration supports exempting companies from consumer data privacy legislation to the extent that their activities are subject to existing Federal data privacy laws. However, activities within such companies that do not fall under an existing data privacy law would be covered by the legislation that the Administration proposes."[124]

There are similar problems with the FTC proposal arising from its sector-based regulation compliance provisions.[125] And, again there is some uncertainty as to how any legislation would apply to information that moves in and out of different regulatory zones.[126] In general terms, these concerns about duplicate burdens are unwarranted in the case of health care regulation. As already discussed, HIPAA/HITECH employs a sector-based confidentiality (disclosure-centric) model. The White House and to an extent the FTC proposals are primarily privacy (collection-centric) endorse models.

---

[119] *See* 42 U.S.C.A. § 17932 (West 2010) (corresponds to HITECH § 13402 and the regulation made thereunder, Breach Notification for Unsecured Protected Health Information, 74 Fed. Reg. 42740-01 (Aug. 24, 2009)).

[120] *See* 42 U.S.C.A. § 17937 (corresponds to HITECH § 13407 and FTC regulation made thereunder, Health Breach Notification Rule, 74 Fed. Reg. 42962-01 (Aug. 25, 2009)).

[121] *See generally* OFFICE OF THE NAT'L COORDINATOR FOR HEALTH INFO. TECH., ABOUT THE PHR MODEL PRIVACY NOTICE: BACKGROUND, DEVELOPMENT PROCESS, KEY POINTS (2011), *available at* http://www.healthit.gov/sites/default/files/phr-model-privacy-notice-backgrounder-final.pdf.

[122] *Privacy & Security Policy*, HEALTHIT.GOV, http://www.healthit.gov/policy-researchers-implementers/personal-health-record-phr-model-privacy-notice (last visited Jan. 30, 2013).

[123] CONSUMER DATA PRIVACY, *supra* note 85, at i.

[124] *Id.* at 38.

[125] *See, e.g.*, Data Security and Breach Notification Act of 2011, S. 1207, 112th Cong. § 2(a)(3).

[126] *See* FED. TRADE COMM'N, *supra* note 86, at 16. "[T]he framework is meant to encourage best practices and is not intended to conflict with requirements of existing laws and regulations. To the extent that components of the framework exceed, but do not conflict with existing statutory requirements, entities covered by those statutes should view the framework as best practices to promote consumer privacy." *Id.*

Straightening out privacy regulation in the face of disparate sectoral protections is not easy. Take, for example, California's health privacy statute that seems less constricted in its applicability and protective of health information even when it exists outside of traditional health care custodians.[127] It does not achieve this feat by protecting all *medical information*.[128] Rather, its ability to regulate, say, a software and services company's PHR platform stems from a deeming provision, which states "[a]ny business organized for the purpose of maintaining medical information in order to make the information available to an individual or to a provider of health care at the request of the individual or a provider of health care, for purposes of allowing the individual to manage his or her information, or for the diagnosis and treatment of the individual, shall be deemed to be a provider of health care subject to the requirements of this part."[129]

However, in the case of federal government protection of health information in the age of big data, these issues can be navigated. HIPAA's weakness here (the fact that it provides only a confidentiality model of protection) can be cast as a strength when it comes to compatibility with the White House and FTC collection-centric models of protection. The sector-based limitations in the proposals should be removed (at least so far as the health sector is concerned) allowing for true *collection* regulation, leaving HIPAA/HITECH to regulate the *disclosure* practices of covered entities. New privacy rules common to all sectors and limiting data collection would then sit upstream of existing health care regulation that would continue to deal with unauthorized information disclosure.

VI. HEALTH RESEARCH CHOICE ARCHITECTURE

The third battleground involving big data and patient privacy is of a different character from the collection-regulation issues discussed in the two preceding sections. The use of health (primarily clinical) data for research is a disclosure issue and one that is currently, but unsatisfactorily, regulated by the HIPAA/HITECH code.

The current HIPAA-research interface is complex. First, as noted in other contexts above, similar or identical activities may be undertaken inside and outside the HIPAA regulatory zone. After all, HIPAA/HITECH only applies to covered entities and so will not apply to data held by many independent researchers or, say, pharmaceutical companies. Further, HIPAA/HITECH

---

[127] CAL. CIV. CODE pt. 2.6 (West 2012).

[128] Rather the definition of "medical information" is derived from health provider source, "any individually identifiable information, in electronic or physical form, in possession of or derived from a provider of health care, health care service plan, pharmaceutical company, or contractor regarding a patient's medical history, mental or physical condition, or treatment." CAL. CIV. CODE § 56.05(g) (West 2012).

[129] § 56.06(a).

operates independently of the Common Rule that might apply to some research by some researchers.[130]

Second HIPAA/HITECH is known for its function creep and multiple exceptions. As a result, a big data-like project such as running data analytics against a hospital's EMR data looking for disease predictors may not be "research" but exempted from regulation as a quality improvement under "health care operations."[131]

Third, HIPAA/HITECH only protects "individually identifiable health information" (and then only in certain circumstances).[132] Data that is de-identified (for example by removing eighteen key identifiers) therefore falls outside of the regulatory zone.[133] Covered entities may also use protected health information for research without authorization under the "limited data set" provision that removes many, but not all, common identifiers.[134]

One of the increased privacy risks introduced by big data is that it can be used to re-identify apparently de-identified data. According to Paul Ohm, "[r]eidentification science disrupts the privacy policy landscape by undermining the faith we have placed in anonymization."[135] This fallacy of de-identification "makes PII-focused laws like HIPAA underprotective by exposing the arbitrariness of their intricate categorization and line drawing. Although HIPAA treats eighteen categories of information as especially identifying, it excludes from this list data about patient visits—like hospital name, diagnosis, year of visit, patient's age, and the first three digits of ZIP code—that an adversary with rich outside information can use to defeat anonymity."[136] Once de-identification is exposed as non-protective, researchers (and others) who seek clinical data increasingly will argue that they should be able to presume the consent of data subjects or that they should be able to rely on minimally restrictive consent models such as opt-out. The next administration must look at the clinical data-research interface anew and build a new, privacy-sensitive choice architecture.

---

[130] *See Federal Policy for the Protection of Human Subjects ('Common Rule')*, U.S. DEPT. OF HEALTH & HUM. SERVS., http://www.hhs.gov/ohrp/humansubjects/commonrule/index.html (last visited Jan. 30, 2013).
[131] *See* 45 C.F.R § 164.506 (2013); *see also* Robertson, *supra* note 18.
[132] *See* 45 C.F.R. § 160.103.
[133] *Id.* § 164.514.
[134] *Id.* § 164.514(e).
[135] Paul Ohm, *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization*, 57 UCLA L. REV. 1701, 1704 (2010).
[136] *Id.* at 1740; *cf.* Daniel C. Barth-Jones, *The "Re-Identification" of Governor William Weld's Medical Information: A Critical Re-examination of Health Data Identification Risks and Privacy Protections, Then and Now* (Columbia Univ. Sch. of Pub. Health, Dept. of Epidemiology, Working Paper No. 2076397, 2012), *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2076397.

A. Research and Big Data

The big data-research interface is complicated. Analytics suppliers are looking at research data (and any clinical data that may include) as pools that feed big data. Further, big data analytics frequently are identified as of potentially great importance in improving the quality and efficiency of health care. As to the latter, at least part of the logic suggested is that health care improvement involves adoption of best practices and best practices are often derived from analysis of big data.[137] Potentially, therefore, predictive analytics could improve processes such as tracking hospital assets and supplies,[138] and improving hospital administrative transactions and asset management.[139] Soon we may also be in a position to learn more about health care financing. For example, major health insurers are providing the Health Care Cost Institute ("HCCI") with de-identified data about more than thirty million persons.[140] Analysis of this data showed how rising costs were likely the major driver for decreased utilization of health care.[141]

Far more controversial, however, are proposals to feed clinical data into big data. For example, it is argued that analysis of this data can drive everything from comparative effectiveness research[142] and clinical decision support to outcomes based pricing models, predictive modeling for new drug research, and public health surveillance and response.[143] A 2012 report by Frost & Sullivan projected an almost forty percent annual growth rate in the use of data analytics by hospitals.[144]

Responsible and respected researchers are beginning to enter this space with some extremely interesting projects. For example, the *eMERGE Network* is a federally funded research project that is combining "DNA biorepositories with electronic medical record (EMR) systems for large-scale, high-throughput

---

[137] MANYIKA ET AL., *supra* note 15, at 39.
[138] Roger Foster, *Reducing Healthcare Administrative Inefficiencies With Big Data*, GOV'T HEALTH IT (May 22, 2012), http://www.govhealthit.com/news/reducing-healthcare-inefficiencies-big-data.
[139] *See id.*
[140] HEALTH CARE COST INST., HEALTH CARE COST AND UTILIZATION REPORT 1 (2010), *available at* http://www.healthcostinstitute.org/files/HCCI_HCCUR2010.pdf.
[141] *Id.* at 16; *see also* N.C. Aizenman, *Data Trove May Shed Light on Health-care Uncertainties*, WASH. POST, May 21, 2012, at A3.
[142] *See* Patrick D. Meek, *CER: Separating Fact from Fiction*, DRUG TOPICS (Aug. 15, 2010), http://drugtopics.modernmedicine.com/drugtopics/article/articleDetail.jsp?id=681894.
[143] MANYIKA ET AL., *supra* note 15, at 42-49; Roger Foster, *Tapping Big Data for Early Identification of Preventable Conditions*, GOV'T HEALTH IT (June 04, 2012), http://www.govhealthit.com/news/tapping-big-data-early-identification-preventable-conditions.
[144] Erin McCann, *Data Analytics Poised for Big Growth*, HEALTHCARE IT NEWS (Aug. 16, 2012), http://www.healthcareitnews.com/news/data-analytics-assume-power-health-it. *See generally* Imran Khan, *U.S. Hospital Health Data Analytics Market*, FROST & SULLIVAN (Aug. 7, 2012), http://www.frost.com/c/10046/sublib/display-report.do?id=NA03-01-00-00-00 (discussing key trends and issues affecting health data analytics market through 2016).

genetic research."[145] *SUrveillance, PREvention, and ManagEment of Diabetes Mellitus (SUPREME-DM)* is a multistate diabetes registry and surveillance project based on data from eleven health care systems.[146] Meanwhile, the University of California's *AMPlab* (Algorithms, Machines, and People), funded by the National Science Foundation, is creating software to make sense of big data repositories provided by federal agencies such as the CDC and NIH (including the Cancer Genome Atlas).[147]

Policy choices here seem much more difficult compared to the big data issues raised above. Many (but not all) of these downstream data processors are responsible, government-funded researchers. And the questions are complicated by related questions such as the "public good" argument about health data and health research data,[148] and "open access" research exemplified by the NIH rules on public access[149] and the growing preference for open research data sets in the United Kingdom[150] and the European Union.[151]

As patient data moves from the treatment to the research domain, processes of consent or authorization become infected with the modern problems that besiege researchers and their subjects such as physicians' conflicts of interest or dual roles of physician and researcher.[152] As already discussed, de-identification is becoming increasingly threatened and, as Paul Ohm convincingly argues, "[d]ata can be either useful or perfectly anonymous but

---

[145] *The eMERGE Network*, eMERGE NETWORK, https://www.mc.vanderbilt.edu/victr/dcc/projects/acc/index.php/Main_Page (last visited Feb. 5, 2013).

[146] *Surveillance, PREvention, and ManagEment of Diabetes Mellitus (SUPREME-DM)*, EDM FORUM, http://www.edm-forum.org/publicgrant/About/projectprofiles/supremedm/ (last visited Fed. 5, 2013). *See generally* Gregory A. Nichols et al., *Construction of a Multisite DataLink Using Electronic Health Records for the Identification, Surveillance, Prevention, and Management of Diabetes Mellitus: The SUPREME-DM Project*, 9 PREVENTING CHRONIC DISEASE 110 (2012), *available at* http://dx.doi.org/10.5888/pcd9.110311.

[147] AMPLAB UC BERKELEY, http://amplab.cs.berkeley.edu (last visited Feb. 5, 2013); *see also* David C. Kaelber et al., *Patient Characteristics Associated With Venous Thromboembolic Events: A Cohort Study Using Pooled Electronic Health Record Data*, 19(6) JAMIA 965 (2012), *available at* http://www.amia.org/sites/amia.org/files/JAMIA-2012-Kaelber-amiajnl-2011-000782.pdf (reporting study of de-identified clinical data of 959,030 patients sourced from EHRs in multiple healthcare systems).

[148] *See* ALEX W. GOODBY, LEIGHANNE OLSEN & MICHAEL MCGINNIS, CLINICAL DATA AS THE BASIC STAPLE OF HEALTH LEARNING: CREATING AND PROTECTING A PUBLIC GOOD (2010).

[149] *Revised Policy on Enhancing Public Access to Archived Publications Resulting from NIH-Funded Research*, NAT'L INST. OF HEALTH (NIH), http://grants.nih.gov/grants/guide/notice-files/NOT-OD-08-033.html (last visited Feb. 11, 2013).

[150] Ian Sample, *Free Access to British Scientific Research Within Two Years*, THE GUARDIAN (July 15, 2012 2:53 PM), http://www.guardian.co.uk/science/2012/jul/15/free-access-british-scientific-research.

[151] Paul Ayris, *European Commission Backs Calls for Open Access*, ASS'N OF EUR. RES. LIBR. (July 17, 2012), http://www.libereurope.eu/blog/european-commission-backs-calls-for-open-access.

[152] BEAUCHAMP & CHILDRESS, *supra* note 89, at 316-19.

never both."[153] Therefore, we have to confront the question of patient consent to the research use of his or her clinical data. Even before the rise of big data and the specter of re-identification became reality, the question was whether researchers should acquire patient authorization for the use of even de-identified clinical data. Today, that question should be re-framed from whether to how.

B. Setting Defaults

Questions such as when and how clinical data should be used for research and what kind of role, if any,[154] patients should have in such decisions are quite complex.[155] Equally, however, the FTC seems committed to a new baseline principle that "[c]ompanies should simplify consumer choice."[156]

Given the competing interests involved (autonomy-derived confidentiality rights in clinical data versus the potential for hugely valuable research), policymakers must find the optimal consent model. Specifically, their search will be for an improved choice architecture. As Richard Thaler and Cass Sunstein point out in *Nudge*, "there is no such thing as a 'neutral' design" of choice architecture.[157] In Siva Vaidhyanathan's words, "[s]ettings only help you if you know enough to care about them. Defaults matter all the time"[158] because, as Thaler and Sunstein explain, defaults are "unavoidable in the sense that for any node of a choice architecture system, there must be an associated rule that determines what happens to the decision maker if she does nothing."[159]

It is clear that, across the board, the FTC wants the industry to improve its choice architecture. The agency's 2012 report included the exhortation that "companies should offer the choice at a time and in a context in which the consumer is making a decision about his or her data."[160] Apparently, including medically inflected data circulating outside of HIPAA protection in their understanding of "sensitive" data that required heightened protection, the agency stated that "before collecting such data, companies should first obtain affirmative express consent from consumers."[161]

---

[153] Ohm, *supra* note 135, at 1703-04.
[154] *See, e.g.*, Franklin G. Miller, *Research on Medical Records Without Informed Consent*, 36 J.L. MED. & ETHICS 560, 560-66 (2008).
[155] *See, e.g.*, Lawrence O. Gostin & James G. Hodge, Jr., *Personal Privacy and Common Goods: A Framework for Balancing Under the National Health Information Privacy Rule*, 86 MINN. L. REV. 1439, 1455-56 (2002).
[156] FED. TRADE COMM'N, *supra* note 86, at 35.
[157] RICHARD H. THALER & CASS R. SUNSTEIN, NUDGE: IMPROVING DECISIONS ABOUT HEALTH, WEALTH, AND HAPPINESS 3 (2008).
[158] VAIDHYANATHAN, *supra* note 37, at 114.
[159] THALER & SUNSTEIN, *supra* note 157, at 83.
[160] FED. TRADE COMM'N, *supra* note 86, at 48.
[161] FED. TRADE COMM'N, *supra* note 86, at 59-60.

Calling for express consent may be appropriate, but "consent" processes are often imperfect in the context of parties with unequal different bargaining positions and in informational asymmetry regarding the implications of any such authorization. Equally, researchers are fearful that any choice architecture that makes it easier for data subjects to opt-out will not only lessen the *quantity* of the clinical data it acquires but also the *scientific* value of the data it does acquire (because opt-outs will distort the data).

Thaler and Sunstein dedicate a chapter in *Nudge* to organ donation[162] and this seems a useful analog from which to approach the consent for clinical data question. In examining extant choice architectures for donations, they note two extreme positions, organ-specific "explicit consent" by family members and a state-run policy of "routine removal" (obligatory consent) of organs.[163] They then examine less radical opportunities for setting choice architecture such as "presumed consent" (akin to opt-out) and "mandated choice" (e.g., drivers license check boxes providing for consent or no-consent).[164]

The relatively simple model for organ donation does stumble somewhat in the clinical data context because of variations in the type of potential data processors. A patient may be more willing to provide his or her clinical data to non-profit medical researchers than, say, pharmaceutical companies or commercial aggregators of big data. It may be that an express consent or an opt-in model would be appropriate in that context. On the other hand, for non-profit or, more narrowly, NIH-funded researchers, an opt-out or mandated choice default may be more appropriate. At the very least data subjects should be given a mandated choice model at the time of admission. Such privacy "moments" are important for reinforcing personal autonomy. That is why privacy advocates decried the Bush administration's removal of the HIPAA privacy rule's requirement of patient consent to disclosure for treatment, payment, or health care operations purposes.[165]

Even with agreement over a core choice architecture, considerable technical difficulties lie ahead. For example, the question *what* information any patient choice should apply to. Most choice architectures assume that the choice will give access to all, say, EMR, information relating to the patient. However, as Mark Rothstein points out, future privacy models may well permit "segmentation" of the record whereby patients could exclude sensitive categories of stored information (e.g., genetic, reproductive, or substance abuse) from typical disclosure models.[166] Other problem areas to navigate include the extent of any pre-choice informed consent and whether later revocation of authorization would be permitted.

---

[162] THALER & SUNSTEIN, *supra* note 157, at 175-82.
[163] *Id.* at 176-77.
[164] *Id.* at 177-80.
[165] 45 C.F.R. § 164.506 (2001), *amended by* 45 C.F.R. § 164.506 (2009).
[166] Rothstein, *supra* note 100, at 394-96, 398-99.

Finally, and at the heart of Thaler and Sunstein's work, policymakers must decide whether to influence a data subject's navigation of any choice architecture with a "nudge" that leverages social norms.[167] It is likely that the clinical data privacy-research tension will persist until the research-treatment dichotomy is rendered both transparent and trustworthy and patients understand how they and their families are investors in and beneficiaries of publicly funded research. If researchers can show progress on these fronts such that clinical data based research is promotable in normative terms then researchers may be deserving of a nudge directed at their data subjects.

VII. CONCLUSION

It is likely that the editors of this Symposium expected the health care contribution to address health reform. That it does not, was not a function of Justice Roberts' taxation "save" of the Affordable Care Act ("ACA") back in June 2012.[168] ACA was always a classic piece of mash-up legislation, a hodgepodge of measures primarily designed to increase access by doubling down on a private insurance model. ACA contains provisions that at least someone has championed as likely to help an ailing health care system or strategically composed to attract the votes of specific House or Senate members. Absent the political will to do the right thing (single payer universal coverage), ACA's controversial nature was a function of its existence in today's bankrupt political climate[169] more than that it contained controversial provisions—not that some have not invented (death panels?) or discovered provisions that can be built out as controversial. During the 2012 presidential election campaign, the choice for the incoming president, rhetoric aside,[170] was never about the repeal of, or the doubling down on, the ACA. Rather, it was about rearranging ACA's deckchairs, finding tweaks that can attract the votes. All that the Roberts court did was set the burden of persuasion; it is now firmly allocated to those who would reduce the health care Congress has promised.

---

[167] THALER & SUNSTEIN, *supra* note 157, at 180-82.

[168] Nat'l Fed'n of Indep. Bus. v. Sebelius, 132 S.Ct. 2566 (2012). "The Affordable Care Act's requirement that certain individuals pay a financial penalty for not obtaining health insurance may reasonably be characterized as a tax. Because the Constitution permits such a tax, it is not our role to forbid it, or to pass upon its wisdom or fairness." *Id.* at 2600.

[169] *See* Paul Krugman & Robin Wells, *Getting Away With It*, N.Y. REV. OF BOOKS (July 12, 2012), http://www.nybooks.com/articles/archives/2012/jul/12/getting-away-it/.

[170] *See, e.g.*, Candidate Mitt Romney: "What the court did not do on its last day in session, I will do on my first day as president of the United States." Jeff Zeleny, *G.O.P. Vowing to Take Battle Into November*, N.Y. TIMES, June 29, 2012, at A1. This was tempered some months before the election but after the Republican convention by Governor Romney's statement that he would keep some provisions of ACA. *See* Lisa Lerer & Margaret Talev, *Romney Says He Would Keep Parts of Obama Health Care Law*, BLOOMBERG (Sept. 9, 2012, 11:01 PM), http://www.bloomberg.com/news/2012-09-10/romney-says-he-would-keep-parts-of-obama-health-care-law.html.

However, it is a mistake to view ACA and the future of health care delivery as distant from this review of patient privacy. HITECH was actually the Obama administration's first foray into health care reform.[171] It sought to establish the health information technology ("HIT") base for key aspects of ACA, for example, models for improving quality and reducing cost such as outcomes and effectiveness research and adverse event reporting.[172] Furthermore, new cost-reducing initiatives such as the Medicare Shared Savings Program that is available to providers who organize as Accountable Care Organizations[173] and the coordinated care model known as the Patient-Centered Medical Home[174] make extensive use of HIT.[175] As EMRs and other HIT initiatives continue to generate vast pools of patient data and data analysis is hyped as the savior of health care, the necessity for a reformed privacy model will increase.

There has been a positive spirit of bipartisanship about HIT and health privacy over the past few administrations and the privacy of health information continues to be broadly embraced by a diverse group of stakeholders. Going into the November 2012 election, both the Democratic and Republican party platforms endorsed privacy (and particularly Internet privacy). The Democratic platform included the White House's proposed Privacy Bill of Rights discussed

---

[171] According to then ONC coordinator Dr. David Blumenthal, "The passage of the Patient Protection and Affordable Care Act of 2010 marks a new era in American health care. Yet in many ways, this era began more than a year earlier, with the passage of . . . [HITECH]." Melinda Beewkes Buntin et al., *Health Information Technology: Laying the Infrastructure for National Health Reform*, 29(6) HEALTH AFF. 1214 (2010), *available at* http://www.gohelp.wv.gov/AdvisoryCouncil/Meetings/Documents/HealthAffairs_June_2010.pdf.

[172] *See generally* Nicolas P. Terry, *Certification and Meaningful Use: Reframing Adoption of Electronic Health Records as a Quality Imperative*, 8 IND. HEALTH L. REV. 45-68 (2011) (describing key provisions of the HITECH Act and its effect on electronic health records).

[173] *See Accountable Care Organizations (ACO)*, CENTERS FOR MEDICARE & MEDICAID SERVS., https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/ACO/index.html?redirect=/ACO/ (last visited Feb. 5, 2013); Ron Klar, *ACO 101: The Basics of Accountable Care*, HEALTH AFF. BLOG (Aug. 29, 2011), http://healthaffairs.org/blog/2011/08/29/aco-101-the-basics-of-accountable-care/.

[174] *See Understanding the Patient-Centered Medical Home*, AM. C. OF PHYSICIANS, http://www.acponline.org/running_practice/pcmh/understanding/index.html (last visited Feb. 5, 2013); Judith Fifield et al., *A Randomized, Controlled Trial of Implementing the Patient-Centered Medical Home Model in Solo and Small Practices*, J. OF GENERAL INTERNAL MED. (2012), *available at* http://link.springer.com/article/10.1007%2Fs11606-012-2197-z; *see also* David W. Bates & Asaf Bitton, *The Future of Health Information Technology in the Patient-Centered Medical Home*, 29(4) HEALTH AFF. 614 (2010), *available at* http://www.cimit.org/images/events/ciw/IT-in-Patient-Centered-Medical-Home.pdf.

[175] *See* Michael Hickins, *Supreme Court Reboots Health IT*, CIO JOURNAL (June 28, 2012, 10:14 AM), http://blogs.wsj.com/cio/2012/06/28/supreme-court-reboots-health-it/?mod=wsj_streaming_supreme-court-health-law-decision; Ken Terry, *Health IT Leaders Divided on Supreme Court Decision*, INFORMATIONWEEK: HEALTHCARE (June 28, 2012, 4:35 PM), http://www.informationweek.com/healthcare/policy/health-it-leaders-divided-on-supreme-cou/240002934.

above.[176] However, the Republican platform, although promising to "ensure that personal data receives full constitutional protection from government overreach and that individuals retain the right to control the use of their data by third parties," made clear the belief that "*the only way to safeguard or improve these systems is through the private sector*."[177] The Democratic victory and the reelection of President Obama (achieved in part by leveraging big data[178]) at least may have reduced the likelihood of such a private sector model. On the other hand, the privacy protection promised by the Democratic platform has not yet been concretized in any legislative proposal. Worse, the corporate-state surveillance complex has been reinforced by President Obama's extension of warrantless electronic surveillance for another five years with his signature of the Foreign Intelligence Surveillance Act.[179] The imminent data takeover requires bipartisan reaction and urgent legislative action that will truly protect patients and their personal health information in the age of big data.

---

[176] *See Moving America Forward*, *2012 Democratic National Platform*, DEMOCRATS.ORG, at 45, *available at* http://assets.dstatic.org/dnc-platform/2012-National-Platform.pdf (last visited Feb. 5, 2013).

[177] *We Believe in America, Republican Platform 2012*, GOP.COM, at 23-24, http://www.gop.com/wp-content/uploads/2012/08/2012GOPPlatform.pdf (emphasis added) (last visited Feb. 5, 2013); *see* Brendan Sasso, *Dem Platform Vows to Defend Internet Freedom*, THE HILL (Sept. 4, 2012, 1:06 PM), http://thehill.com/blogs/hillicon-valley/technology/247381-democrats-vow-to-defend-internet-freedom-in-platform

[178] Michael Scherer, *Inside the Secret World of the Data Crunchers Who Helped Obama Win*, TIME (Nov. 07, 2012), http://swampland.time.com/2012/11/07/inside-the-secret-world-of-quants-and-data-crunchers-who-helped-obama-win/#ixzz2KcC0yheE.

[179] Associate Press, *Obama Signs FISA Warrantless Wiretapping Program Extension Into Law*, HUFFINGTON POST (Dec. 30, 2012, 5:49 PM), http://www.huffingtonpost.com/2012/12/30/obama-fisa-warrantless-wiretapping_n_2385690.html.