RISK-BASED REGULATION OF CLINICAL HEALTH DATA ANALYTICS

DEVEN MCGRAW & ALICE LEITER*

INTRODUCTION

The United States is undergoing a welcome but long overdue digital health care revolution. Due at least in part to taxpayer-funded financial incentives mandated by Congress in the 2009 Health Information Technology for Economic and Clinical Health Act (HITECH),[1] the percentage of physicians using an advanced electronic health record

* Deven McGraw is a partner at Manatt, Phelps & Phillips LLP and Alice Leiter is an associate at Hogan Lovells. At the time this paper was written, Deven McGraw was the Director, and Alice Leiter was Policy Counsel, for the Health Privacy Project at the Center for Democracy & Technology (CDT) in Washington, D.C.

1. Health Information Technology for Economic and Clinical Health Act, Pub. L. No. 111-5 § 13001, 123 Stat. 226 (2009).

(EHR) system has almost tripled in the last five years; for hospitals, such use has skyrocketed from roughly nine percent in 2008 to more than 80 percent in 2013.[2] The benefits of this increase in EHR adoption are already being realized: recent studies show that 94 percent of health care providers report that their EHRs make patients' records available at the point of care; 88 percent report that their EHR produces clinical benefits for their practice; and 75 percent report that their EHR has improved the quality of the patient care they are able to deliver.[3]

Patients are also increasingly using the Internet and mobile tools to collect and share personal information relevant to their health and well-being; it is estimated that there are over 40,000 mobile health applications across multiple platforms and that 247 million people have downloaded a health app.[4] The merger of these two worlds—the traditional health information ecosystem, historically dominated by medical care delivery (and payment) settings, and the patient-facing ecosystem—is well underway. Beginning in 2014, providers participating in the federal EHR incentive program are required to provide their patients with digital access to downloadable and sharable clinical data relevant to their health, such as test results, medical record copies and family health history.[5] The Department of Health and Human Services (HHS) is considering proposals to reward providers for accepting digital data from patients, and a number of forward-thinking health care organizations have already begun to implement care models that involve integration of health information submitted by patients.[6]

Supporters of initiatives to digitize medical information also hope to leverage clinical patient data to glean better, faster insights into which types of treatments and prevention strategies work best and in which particular subpopulations. It is well-known that those in the U.S. pay more for care than anywhere else in the world, and yet their health outcomes significantly lag as compared to those in other countries.[7]

2. Press Release, U.S. Department of Health & Human Services, Doctors and Hospitals' Use of Health IT More than Doubles Since 2012 (May 22, 2013), *available at* http://www.hhs.gov/news/press/2013pres/05/20130522a.html.

3. Office of the Nat'l Coordinator for Health IT, *Improved Diagnostics and Patient Outcomes,* HEALTHIT.GOV, http://www.healthit.gov/providers-professionals/improved-diagnostics-patient-outcomes (last visited Feb 17, 2014).

4. Darrell West, *How Mobile Devices are Transforming Healthcare*, ISSUES IN TECHNOLOGY INNOVATION (May 2012), *available at* http://www.brookings.edu/~/media/research/files/papers/2012/5/22%20mobile%20health%20west/22%20mobile%20health%20west.pdf.

5. 45 C.F.R. § 170.314(e)(1) (2014).

6. *See* CENTER FOR CONNECTED HEALTH, http://www.connected-health.org/ (last visited Mar. 27, 2014); *see also* PROJECT HEALTH DESIGN, http://www.projecthealthdesign.org (last visited Mar. 27, 2014).

7. Gerard F. Anderson, Uwe E. Reinhardt, Peter S. Hussey & Varduhi Petrosyan, *It's the*

Further, there are troubling disparities with respect to the care received by Americans of particular racial or ethnic backgrounds.[8]

To gain full value from the significant public investment in the use of EHRs, the U.S. health system needs to more robustly leverage health data initially collected in medical records for further analytic or "learning" purposes. To assure public trust in the secondary use (or re-use) of digital medical information for learning purposes, patient privacy concerns will need to be effectively addressed.

This article explores the potential harms from misuse or inappropriate use of medical information; the ways that current federal regulations attempt to address risks of harm; and why current rules governing re-use of medical information for analytic purposes are not sufficiently targeted to minimizing such risks. It argues that a more effective legal framework would be one in which protections and restrictions are commensurate with the "riskiness" (potential for harm) of the data use: the greater the risk and potential for harm posed by a particular type of data activity, the greater the protections should be. Along those same lines, when the risk and potential for harm are low, this paper urges that data can and should be used and exchanged with greater ease and flexibility than is currently possible. The article closes by suggesting data characteristics and data-sharing activities that arguably increase the risk of harm, in the hope of laying the foundation for a more risk- (or harm-) based approach to regulating health data analytics.

The article focuses specifically on policy frameworks to protect the privacy of health information collected within the traditional health care system and utilized for learning purposes, though it recognizes the additional privacy risks facing consumers who are increasingly sharing their health information in spaces not regulated by comprehensive health privacy laws. This latter type of data holds great value for analytic purposes; however, this article is intentionally focused on health data collected in *clinical* settings.[9]

The article uses a specific example of a regulation that currently is not sufficiently based on the risk of harm in order to begin exploring how it and other regulations should evolve—through application of the fair

Prices, Stupid: Why the United States is So Different From Other Countries, 22 HEALTH AFF. 89 (2003).

8. Ichiro Kawachi, Norman Daniels & Dean E. Robinson, *Health Disparities by Race and Class: Why Both Matter*, 24 HEALTH AFF. 343 (2005).

9. We note that this analysis is confined to the data collected from health care provider records, as opposed to that from both providers and payers. Although claims data can be quite valuable for analytic purposes, our intent is to start by reworking the framework for provide, which will give us a foundation for thinking through a similar revised framework for payer data.

information practice principles (FIPPs)—to more appropriately fit the current digital health environment. The hope is that this admittedly narrow lens will lay some early groundwork for thinking about health privacy policy frameworks for the increasingly rich realm of health data outside of the coverage of HIPAA.

WHAT ARE HEALTH INFORMATION PRIVACY HARMS?

What is unique about health data?

Health information—particularly the type of health information collected by physicians and hospitals in clinical care settings—is generally agreed to be among the most sensitive categories of personal information. A medical record often contains details about an individual's most basic biologic makeup, and from its contents one could learn a range of intimate information about a person's life, extending even to inferences about the health status of family members. In survey data, consumers and patients consistently express concerns about the privacy and confidentiality of their health information beyond those they have about non-health information.[10]

Privacy legal regimes typically include special protections for health or medical information. In HIPAA, Congress tasked HHS with the responsibility of developing privacy and security regulations to govern identifiable health information used and disclosed by health care providers, health plans, and healthcare clearinghouses. All states have laws governing the use and disclosure of health information, with some placing a greater emphasis—with more comprehensive protections—on medical privacy than others.[11] Recently, reports on the need to more effectively protect consumer privacy (issued in 2012 by both the White House and the Federal Trade Commission (FTC)) identify health data as having a level of sensitivity beyond most routine personal information.[12]

Another factor that distinguishes health data from other types of

10. CAL. HEALTHCARE FOUND., NATIONAL CONSUMER HEALTH PRIVACY SURVEY 2005 (November 2005), *available at* http://www.chcf.org/publications/2005/11/national-consumer-health-privacy-survey-2005; NAT'L P'SHIP FOR WOMEN AND FAMILIES, MAKING IT MEANINGFUL: HOW CONSUMERS VALUE AND TRUST HEALTH IT (February 2012), *available at* http://go.nationalpartnership.org/site/PageServer?pagename=issues_health_IT_survey.

11. *See, e.g.*, California Confidentiality of Medical Information Act, CAL. CIV. CODE §§ 56-56.16 (2013); New York State Public Health Law, N.Y. PUB. HEALTH LAW §§ 17-18 (McKinney 2013).

12. FED. TRADE COMM'N, PROTECTING CONSUMER PRIVACY IN AN ERA OF RAPID CHANGE: RECOMMENDATIONS FOR BUSINESSES AND POLICYMAKERS, 47 (2012); WHITE HOUSE, CONSUMER DATA PRIVACY IN A NETWORKED WORLD: A FRAMEWORK FOR PROTECTING PRIVACY AND PROMOTING INNOVATION IN THE GLOBAL DIGITAL ECONOMY, 11 (2012).

personal data is its potential to contribute to the social good, especially when used collectively. As noted earlier, health system improvement efforts depend upon the use and reuse of data originally collected in medical records. In general, laws governing personal health information specifically allow for uses of health data—often without the need for the specific consent or authorization of the data subject—for public health (such as disease surveillance), for health improvement initiatives, and to meet other public policy needs, such as law enforcement. Consequently, laws intended to protect the confidentiality of health data, enacted in response to its level of sensitivity, also need to accommodate uses of that data that contribute to the common good.

Potential Health Privacy Harms

As explored and discussed by privacy scholar Ryan Calo, harms resulting from a breach of privacy can be both subjective—referring to the perception of unwanted observation, resulting in unwelcome mental states; and objective—the unanticipated or coerced use of information concerning a person against that person.[13] Examples of subjective privacy harms include, but are not limited to: discrimination in any area of one's life; damage to one's reputation, whether real or perceived; or any form of embarrassment. Objective harms include: financial harm; physical harm; or the theft of one's identity.[14]

Within a health context, subjective harms might include the experience of being treated differently by peers because of a known chronic medical condition, being ashamed that neighbors or colleagues know about a substance abuse issue, or being shunned because of a mental health condition.[15] Objective harms might include losing a job or an opportunity for a new position because an employer knows of a health condition, or—prior to the implementation of the Affordable Care Act[16]—being discriminated against for purposes of procuring health insurance. Other types of insurance—including life and disability—frequently use health status in determining the extent of coverage and the cost.[17]

13. Ryan Calo, *The Boundaries of Privacy Harm*, 86 INDIANA L.J. 1131, 1142 (2010).

14. *Id.* at 1143.

15. *See, e.g.,* Graham Thornicroft, Diana Rose & Aliya Kassam, *Discrimination in Health Care Against People with Mental Illness*, 19 INT'L REV. OF PSYCHIATRY 113 (2007).

16. Patient Protection and Affordable Care Act, Pub. L. No. 111-148, 124 Stat. 119 (2010).

17. *See*, *e.g.*, *Benefits for People with Disabilities*, SOCIAL SECURITY ADMINISTRATION, http://www.ssa.gov/disability/ (last visited Feb. 17, 2014); *see also Understanding Life Insurance*, TEX. DEP'T OF INS., http://www.tdi.texas.gov/pubs/consumer/cb018.html (last visited Feb. 17, 2014); *How Life Insurance Rates are Determined*, INSURE.COM, http://www.insure.com/articles/lifeinsurance/underwriting-categories.html (last visited Feb.

Harm to individuals' trust in the health care system also plays a crucial role with respect to possible consequences of failure to protect health privacy. Surveys have shown that individuals do not seek care, withhold information or lie about their medical conditions if they do not trust that the information will be kept confidential. In a recent survey, one out of eight individuals admitted that he had withheld health information from their providers because of concern about that information's security or safety.[18]

This is a form of subjective harm, since it creates a mental state of apprehension and mistrust—but with a unique twist. Based on the number of individuals who have admitted to engaging in privacy-protective behaviors compared to the number of individuals who have themselves experienced an actual health data breach, it appears that not all of those who admit to withholding health information or misrepresenting their health history have ever personally experienced a breach. Surveys tend to show that a smaller number of individuals report having experienced a breach of their health information than do those reporting taking privacy-protective behaviors; a 2011 study put the former number at one in 25.[19] Thus, it seems likely that this "harm to trust" can occur even among patients who themselves have never had—or knowingly had—their data breached or misused.

This lack of trust in the health care system has real implications for both individual and population health. An individual who does not seek treatment—or who lies about her condition—is far less likely to obtain appropriate care. And a health care system in which some portion of the data is inaccurate or incomplete is less likely to itself generate valuable analytics or population health improvements. This is of particular concern given that privacy worries are expressed more consistently by racial and ethnic minorities than any other population subgroup, and that they also continue to suffer from disparities in care.[20]

17, 2014).

18. Israel T. Agaku, Akinyele O. Adisa, Olalekan A. Ayo-Yusuf, & Gregory N. Connolly, *Concern about Security and Privacy, and Perceived Control over Collection and Use of Health Information are Related to Withholding of Health Information from Healthcare Providers*, 21 J. AM. MED. INFORM. ASSOC. 374 (2014).

19. NAT'L P'SHIP FOR WOMEN AND FAMILIES, *supra* note 10.

20. Michael V. Laric, Dennis A. Pitta & Lea Prevel Katsanis, *Consumer Concerns for Healthcare Information Privacy: A Comparison of U.S. and Canadian Perspectives*, 12 RES. IN HEALTHCARE FIN. MGMT. 93 (2009).

FEDERAL REGULATION OF RISKS TO HEALTH INFORMATION[21]

When it comes to collection, use, and reuse of health information, the most relevant federal law is HIPAA. As noted above, HIPAA regulates health information only when collected, used, and disclosed by health care providers,[22] health care insurers, and health care clearinghouses (and contractors acting on their behalf);[23] it does not apply to all health information.

The details of HIPAA's privacy and security protections are found in its regulations. For the most part, the rules are designed to minimize risk of harm. For example, the Privacy Rule applies only to individually identifiable health information. Data that is "de-identified"—and that raises very low risk of re-identification—is not subject to any regulation as long as such de-identification is done using an approach recognized in the Privacy Rule.[24] The Privacy Rule also allows data that has been stripped of some common identifiers—and is thus less risky—to be used for research, public health, and administrative operations without the need to obtain prior specific patient authorization.[25] Because there is still some risk of re-identification, recipients of this "limited data set" are required to execute data use agreements, which must include commitments not to re-identify the data.[26]

Another way HIPAA regulates risk is by allowing most routine uses of identifiable health information without the need to first obtain consent of the data subject. However, the law requires fairly specific authorization for data use activities that arguably are not routine, or may not be expected by data subjects. Uses for purposes of "treatment, payment and health care operations" are routine, but those for research

21. We note that an examination of state laws regulating risks to health information is beyond the scope of this paper. All states do contain protections for the use and disclosure of health information and vary with respect to how stringent such protections are.

22. Only health care providers who bill for services electronically using standard HIPAA transaction code sets are covered by HIPAA.

23. A "health care clearinghouse" is a "public or private entity, including a billing service, repricing company, community health management information system or community health information system, and 'value-added' networks and switches, that does either of the following functions: (1) Processes or facilitates the processing of health information received from another entity in a nonstandard format or containing nonstandard data content into standard data elements or a standard transaction; or (2) Receives a standard transaction from another entity and processes or facilitates the processing of health information into nonstandard format or nonstandard data content for the receiving entity." 45 C.F.R. § 160.103 (2013).

24. There are two acceptable methods under HIPAA to de-identify health data. One approach is the "Safe Harbor" method that removes the 18 categories of identifiers. The other approach is the statistical method that uses expert statistical analysis to achieve "very small" chance of re-identifying the data. 45 C.F.R. § 164.514 (2013).

25. 45 C.F.R. § 164.514 (2014).

26. *Id.*

are not; thus, under HIPAA, research requires individual data subject authorization, unless the entity has obtained a waiver of such requirement after a review by an Institutional Review Board (IRB) or Privacy Board—committees that have been designated to review and monitor biomedical research involving human subjects.[27]

When data are used by "business associates"—entities that have entered into contractual arrangements with health care providers or insurers to store or use data for particular purposes—they are required to execute enforceable agreements setting forth their specific rights and obligations with respect to the data.[28]

Other federal rules apply to specific types of health information, or to health information in particular circumstances. For example, the rules governing federally funded substance abuse treatment programs presume the data collected by these programs raises heightened risk of harm. Those regulations require specific authorization from the patient before information identifying the patient as a potential substance abuser may be disclosed to third parties.[29]

The Common Rule regulates research using identifiable health information that is conducted with the support of federal funding from certain agencies, including, among others, the Department of Health and Human Services, the Department of Veterans Affairs and the Department of Energy. Consistent with how HIPAA treats uses of data for research, the Common Rule also requires the approval of an Institutional Review Board (IRB) for research uses of health information, and the prior authorization of the data subject.[30]

WHY THE CURRENT REGULATORY FRAMEWORK FOR RE-USES OF HEALTH DATA IS NOT SUFFICIENTLY RISK-BASED

As explained in more detail below, HIPAA's rules governing re-use of identifiable health information for learning purposes do little to reduce the risk of harm to patients and instead create disincentives to share the

27. *See* 45 C.F.R. § 512(i) (2014). The authorization requirement can be waived if the IRB or Privacy Board finds that the following criteria have been met: (1) The use or disclosure of protected health information involves no more than a minimal risk to the privacy of individuals; (2) the research could not practicably be conducted without the waiver or alteration; and (3) the research could not practicably be conducted without access to and use of the protected health information.

28. 45 C.F.R. § 164.308(b)(3) (2013).

29. 42 U.S.C. §§ 290dd-2(a)-(b) (1998).

30. Under the Common Rule, informed consent can be waived if the IRB finds and documents that the research involves no more than a minimal risk to the subject(s), that a waiver will not adversely affect the rights and welfare of the subject(s), that the research could not be practically carried out without a waiver, and that the subjects will be provided with additional pertinent information after participation. 45 C.F.R. § 46.116(d) (2001).

results of health data analytics for learning purposes.

Paradox

The HIPAA Privacy Rule presumes that the risk of harm to patients is greater from research uses of data than uses of health data for "operations." Under current law, "health care operations" includes:

- "Conducting quality assessment and improvement activities, including outcomes evaluation and development of clinical guidelines, provided that the obtaining of generalizable knowledge is not the primary purpose of any studies resulting from such activities;" and
- "Population-based activities relating to improving health or reducing health care costs, [and] protocol development. . ."[31]

"Research" is defined as a "systematic investigation, including research development, testing, and evaluation, designed to develop or contribute to generalizable knowledge."[32] The Common Rule uses the same definition for "research."

Consequently, if an analysis of health information for quality improvement or population health purposes is not intended to produce results that will be shared with others for learning purposes, it is treated as "operations," and it can be conducted under the general oversight of the entity covered by the law, without the need for prior consent from the data subjects or approval of a Privacy Board or an IRB. Because such uses are not "research," they would not be regulated by the Common Rule.

The paradox of these provisions is that two studies analyzing the same data for quality improvement purposes, exposing the same data points, to address the same question or sets of questions, and done by the same institution, will be treated as operations if the results are only to be used internally; and treated as research if the intent is to share the results with others. In circumstances where the release of the results does not itself raise risks to privacy (because the results are in de-identified, aggregate form), this distinction neither advances the learning health system nor reduces risk to patient privacy or potential for harm to the data subjects.

This paradox is reinforced by cautious implementation. Because the decision to share the results of a quality improvement or other internal study is not always made until after the analysis has been conducted,

31. 45 C.F.R. § 164.501 (2014).
32. *Id.*

entities tend to play it safe and treat quality analytics and population health projects uniformly as research.[33] The strict research requirements give entities incentive to jump through all necessary hoops on the front end, in order to avoid having to comply retroactively; on the other end of the spectrum, they also may contribute to reluctance on the part of health care entities to share the results of their QI activities externally.

RETHINKING THE REGULATORY FRAMEWORK

Ideally, the regulation of the collection and use of health information for learning purposes should be based on risk of harm. In the case of health information, a definition of harm should include the potential damage to patient trust that can occur when there is lack of confidence that the information will be protected and not used inappropriately or detrimentally.

HIPAA and the Common Rule both presume that the publication of study results—to contribute to "generalizable knowledge"—renders any re-use of health information for learning purposes more risky. But any risks that derive from publication of results can be accommodated—for example, by requiring that results be published in de-identified form.[34] Current requirements do not address any risks that derive from the analysis of the underlying, typically patient-identifiable health data. The mere intent to publish or share what is learned should not, by itself, trigger more robust regulation assuming sufficient protection against re-identification. And, if a particular publication is shown to raise risk for a specific reason, that increased risk should itself trigger more regulatory safeguards.

Below we assess what aspects of health information re-use for analytic purposes raise greater risk of harm and what types of protections should be required in order to better protect against that risk. We argue that a sliding scale of protections, based on these risk factors, is more appropriate for analytic uses of clinical data for learning purposes.

33. Deven McGraw & Alice B. Leiter, *Pathways to Success for Multi-Site Clinical Data Research*, 1 EGEMS (GENERATING EVIDENCE & METHODS TO IMPROVE PATIENT OUTCOMES), Iss. 1, no. 13, Sept. 19, 2013, at 8.

34. De-identified data is not regulated by either HIPAA or the Common Rule. Nevertheless, de-identification does not result in zero risk of re-identification. Deven McGraw, *Building Public Trust in Uses of Health Insurance Portability and Accountability Act De-identified Data,* 20 J. AM. MED. INFO. ASS'N. 29, 30 (2012). Consequently, risk-based regimes for regulating health data analytics may need to include additional safeguards to address even the residual risk of re-identification.

What Raises the Risk of Privacy Harm?

Internal vs. External

A more reasonable, nimble, and goal-oriented regulatory framework would treat the use of clinical data to evaluate safety, quality, and efficacy in the same way as operations are treated, even if the intent is to share results for generalizable knowledge, as long as the provider entity maintains sufficient oversight and control over data use decisions. HIPAA's more relaxed regulatory treatment of health care operations presumes that "internal" uses of health information for analytic purposes are more routine and do not heighten risks to privacy—likely because survey data shows that, in general, patients tend to trust their health care providers with respect to the confidentiality of their health information.[35]

Should this "internal use" designation be limited only to those circumstances where the raw, patient-level data does not leave the physical confines of the organization, or is it possible to consider a use to be "internal" if the organization has sufficient contractual or other controls over uses of the information? The Privacy Rule permits healthcare providers to share patient data with other entities covered by HIPAA for operations purposes (including the quality analytics that are the focus of this article) for patients that they have in common. In addition, current rules clearly favor a definition of "internal" that allows for some external sharing under contractual controls.[36] As noted above, the Privacy Rule permits providers to hire contractors or "business associates" to perform contractually specified services on their behalf.

Questions have been raised about whether these business associate agreements (BAAs) provide sufficient protections for patients. Frequently, the economic bargaining power of a business associate could be greater than that of the provider organization seeking its services. Consequently, the BAA may be written in a way that permits a fairly wide berth of health information uses, subject to the outer boundaries of what is permitted by law.[37] Such agreements further raise the risk to data

35. *See, e.g.*, NAT'L P'SHIP FOR WOMEN AND FAMILIES, *supra* note 10. This same level of trust does not tend to exist between consumers and their health insurers; almost 50 percent of surveyed individuals say they do not trust their insurer. MEASURING THE VALUE OF TRUST IN HEALTHCARE 3, (Peppers & Rogers Group, 2012), *available at* https://www.worldcongress.com/events/HW12084/pdf/WPPRG_TrustinHealthcare.pdf.

36. This approach is one that has been recommended by the federal Health IT Policy Committee. Office of the Nat'l Coordinator for Health IT, *Recommendations to the National Coordinator for Health IT,* HEALTHIT.GOV, http://www.healthit.gov/facas/health-it-policy-committee/health-it-policy-committee-recommendations-national-coordinator-health-it (last visited Oct. 7, 2014).

37. The federal Health IT Policy Committee laid out its concerns on this front in a September 2010 transmittal letter to the National Coordinator for Health Information

if they contain no explicit—or weak—limits on the length of retention of patient information, or if they lack sufficient requirements to return or destroy data once the particular need for data has expired.

In addition, although Congress granted authority to federal and state regulators to hold business associates directly accountable for compliance with HIPAA,[38] compliance with more stringent limitations on data use in BAAs is more likely to be the obligation of the contracting provider. Large health care providers frequently have hundreds of BAAs, rendering it difficult (if not impossible) to establish high expectations for enforcement.

Further consideration of "internal vs. external" issues should include patients' reasonable expectations regarding uses and disclosures of their health information,[39] as well as whether the entities receiving the data are subject to HIPAA or some other form of public accountability. Further thought should also include finding ways to create incentives for data sharing structures that minimize risk to data privacy and security, such as "decentralized" or "federated" research networks. Such structures allow multiple institutions to share study results without physically moving the underlying data, as the analytics are conducted at each collecting institute and then aggregate results are later combined.

Level of Sensitivity of the Data

A more risk-based regulatory framework for analytics could require heightened protections for studies involving identifiable sensitive information. The Privacy Rule provides uniform protections to all types of health information, with the exception of psychotherapy notes, which are afforded enhanced protections.[40] However, other privacy laws provide certain types of health data with greater protections. As noted above, identifiable health data from federally funded substance abuse treatment programs are subject to heightened protections,[41] and genetic

Technology. Letter from Paul Tang, Vice Chair, Health IT Policy Comm., to David Blumenthal, Nat'l Coordinator for Health Info. Tech., U.S. Dep't of Health & Human Services (Sept. 1, 2010), *available at* http://www.healthit.gov/sites/default/files/hitpc_transmittal_p_s_tt_9_1_10_0.pdf.

38. 45 C.F.R. § 160.102(b) (2013).

39. *See, e.g.*, HELEN NISSENBAUM, PRIVACY IN CONTEXT: TECHNOLOGY, POLICY, AND THE INTEGRITY OF SOCIAL LIFE 89–98 (2010); *see also* Andrew Selbst, *Contextual Expectations of Privacy*, 35 CARDOZO L. REV. 643 (2013). We note also that the Health IT Policy Committee established a fundamental principle that the patient should not be "surprised" by what happens to their data, though of course there may be limits to this principle as patients may in fact be fairly uninformed about even routine uses of their data. Letter from Paul Tang, *supra* note 37, at 4.

40. 45 C.F.R. § 164.501 (2013); 45 C.F.R. § 164.508(a)(2) (2013).

41. 42 U.S.C. §§ 290dd-2(a)-(b) (1998).

information is protected by federal law against its use to discriminate in employment and health insurance.[42] Many states have moved to protect categories of sensitive data more stringently than non-sensitive data, requiring additional levels of notice and/or consent when these types are used or disclosed.[43] State laws also frequently provide minors with rights to privacy with respect to certain types of health data.[44]

The National Committee on Vital and Health Statistics (NCVHS), a federal advisory committee, has published a number of papers recommending various categories of data be considered sensitive and thus deserving of special consideration, most recently specifying: genetic information, psychotherapy notes, substance abuse treatment records, HIV and other sexually transmitted disease information, information in the records of children and adolescents, mental health information (beyond that contained in psychotherapy notes), and sexuality and reproductive health information.

Development of more risk-based policies for analytics should also consider whether heightened protections are needed for vulnerable populations. Studies have shown that minorities and low-income populations tend to be among the least trusting of research uses of their data.[45] Further, persons who are very ill, when surveyed, tend to ascribe a very low value to privacy, making them especially vulnerable to misuse of data.[46] However, since these populations are also in the greatest need

42. Genetic Information Nondiscrimination Act of 2008, 42 U.S.C. §§ 2000ff(-1)-(-11) (2008).

43. *See, e.g.*, with respect to HIV-related information: MO. REV. STAT. § 191.656 (2012); HAW. REV. STAT. § 325-101 (2013); N.Y. PUB. HEALTH LAW ARTICLE 27-F; ARIZ. REV. STAT. ANN. § 36-664 (2013).

44. *See, e.g.*, with respect to obtaining outpatient substance abuse treatment, WASH. REV. CODE ANN. § 70.96A.096 (2013).

45. *See, e.g.*, Donald Musa et. al., *Trust in the Health Care System and the Use of Preventive Health Services by Older Black and White Adults*, 99 AM. J. PUB, HEALTH 1293 (2009); Vanessa B. Sheppard et. al., *Providing Health Care to Low-Income Women: A Matter of Trust*, 21 FAM. PRAC. 484 (2004); Lorenzo Moreno et. al., *Personal Health Records: What Do Underserved Consumers Want?*, MATHEMATICA ISSUE BRIEF, May 2007, *available at* http://www.mathematica-mpr.com/publications/pdfs/phrissuebr.pdf.

46. A recent survey of 2,125 members of the online social network PatientsLikeMe, made of adults with health conditions, showed that an overwhelming majority would be willing to share health data if it could help others in some way: 94% would be willing to share to help doctors improve care; 94% would be willing to help other patients like them; and 92% would be willing to share to help researchers learn more about their disease. Four out of five respondents (84%) would be willing to share their health information with drug companies to help them make safer products, and 78% would do so to let drug companies learn more about their disease. 94% believe that their health data *should* be used to improve the care of future patients who may have the same or similar condition. *PatientsLikeMe Survey Shows Vast Majority of People With Health Conditions Are Willing To Share Their Health Data*, PATIENTSLIKEME (Jan. 23, 2014, 10:00 AM), http://news.patientslikeme.com/press-release/patientslikeme-survey-shows-vast-majority-people-health-conditions-are-willing-

of the valuable outcomes of data sharing for learning purposes, the rules established to protect these populations should also enable, encourage, and facilitate more robust analysis.

Failure to Establish and Adhere to FIPPs-Based Policies

The principles of fair information practices, or FIPPs, are the foundation for most privacy laws, including those governing health information, both in the U.S. and internationally. Consequently, failure to adhere to FIPPs with respect to re-uses of health information for learning purposes arguably increases the risks of privacy harm. Although there are many versions of FIPPs, the below analysis relies on the articulation put forth by the Markle Foundation's Common Framework, which was developed by its multi-stakeholder Connecting for Health Steering Group.[47]

The text below represents some initial thoughts about the types of activities with health data raise higher risk.

Openness and Transparency:

The HIPAA Privacy Rule requires, in the interest of transparency, that health care providers give their patients a Notice of Privacy Practices.[48] This Notice is required to include, among other items, the permitted uses and disclosures of a patient's health care information without the need for authorization, and those uses and disclosures that do require patient authorization. The Notice also must specify the rights of patients to access their health information and to request corrections.

It is important for patients to understand their rights and the basic legal regime protecting their health information; however, transparency to patients should also include some information about *actual* uses and disclosures of health information. Although it could be onerous to spell out in detail every single analytic use of patient data (and it is often even more difficult to identify whether any one patient's data was used in a particular analysis), patients should at least be able to obtain easily a summary of types of learning activities to which their data may have contributed. Failure to provide this type of transparency to patients arguably increases the risk of harm, particularly harm to patient trust in the health care system.

share-t.

47. *See Markle Common Framework*, MARKLE FOUNDATION, http://www.markle.org/health/markle-common-framework (last visited Feb. 17, 2014).

48. 45 C.F.R. § 164.520 (2013).

Data Minimization

The HIPAA Privacy Rule requires health care providers to use the minimum amount of health information necessary for the particular purpose for which data is accessed.[49] This "minimum necessary" standard has been part of the Privacy Rule since its inception; in the absence of best practices on how to apply it, Congress in HITECH required HHS to establish guidance on HIPAA's minimum necessary standard no later than August of 2010.[50] To date, no such guidance has been released. Failure to establish controls on how much data is utilized for analytics arguably increases the risk of harm.

As discussed above, the Privacy Rule already establishes incentives to perform analytics with data that is less identifiable. Such information, if sufficiently protected against unauthorized re-identification, arguably raises less risk to privacy and thus its uses should be encouraged wherever appropriate and possible. The concept of data minimization should be applied to the identifiability of the data—but the extent of health information collected to address a particular analytic question should also not go beyond what is reasonably needed to answer the question.

Collection, Use and Disclosure Limitations

How much information is collected for analytic purposes, and to how many people it is exposed (both through collection and disclosure), are also arguably factors in calculating risk of harm.[51] If there are sufficient controls on the amount of information collected and on the persons to whom the information is exposed, even in an analysis where the results are only intended to be used internally, the risk of harm is reduced.

Security Safeguards

Entities covered by HIPAA—including health care providers and their business associates—are required to implement reasonable security safeguards for data in paper format and to abide by the HIPAA Security Rule for electronic identifiable health information. The failure to maintain such safeguards—or the release of health information into

49. 45 C.F.R. §§ 164.502(b), 164.514(d) (2013).

50. 42 U.S.C. § 17935(b)(1)(B) (2009).

51. Indeed, the mere collection of consumer data, due to possibilities of breach, misuse, or unauthorized access, can implicate privacy interests. JUSTIN BROOKMAN & G.S. HANS, WHY COLLECTION MATTERS: SURVEILLANCE AS A DE FACTO PRIVACY HARM (Future of Privacy Forum Big Data & Privacy Workshop Paper Collection, 2013), *available at* http://www.futureofprivacy.org/wp-content/uploads/Brookman-Why-Collection-Matters.pdf.

environments where such safeguards are weak or uncertain—raises the risk of harm.

Accountability and Oversight

Analytic uses of health information that qualify as research under the Privacy Rule trigger fairly specific provisions with respect to accountability, including Privacy Board or IRB review, as discussed above. If the information is fully identifiable, the authorization of the patient is required.[52] As noted above, data that is either de-identified or less identifiable, such as a limited data set, may be used without consent, although limited data set information must be protected by a data use agreement.[53]

In contrast to the accountability and oversight requirements governing research, analytic activities that qualify as health care operations under HIPAA are not subject to particular oversight requirements beyond those that apply to other routine uses of health information. Use of health information for analytic purposes should be subject to some mechanism of oversight, with the degree of rigor dependent on the risks of harm. For example, periodic audits of internal oversight of such activities—even if conducted only internally, and with only aggregate, de-identified results shared—would provide an appropriate measure of accountability.

Characteristics of a Re-imagined Framework

A risk-based framework of protections for analytic uses of clinical health information could include a sliding scale of protections and accountability and oversight requirements that is based on risks of harm, rather than solely on whether or not the results of the analysis will be shared for "generalizable knowledge." Such a framework would reduce specific requirements on research re-using clinical data that:

- Is performed internally or under tight contractual controls;
- Minimizes (both with respect to content and exposure to others) the information that is used for analysis;
- Involves health data not posing additional risk due to the vulnerability of the subjects;
- Is transparent to the public (or at least the community of likely data subjects); and
- Reports results in a way that does not raise additional

52. 45 C.F.R. § 164.512(i) (2013). As noted above, there are certain conditions under which authorization requirements can be waived by an IRB or Privacy Board.

53. 45 C.F.R. § 164.514(e) (2013).

privacy risks.

Additional safeguards—including requiring patient authorization for research triggering the most risk of harm—would be applied to analytics that do not meet the characteristics of lower risk.

CONCLUSION

More clearly defining a risk-based framework for governing analytic uses of health data has the potential to enable analytics that could—and should—be conducted easily, but currently are not, due to uncertainty about application of current law and/or typical risk-averse behavior among health care entities and researchers. In addition, specifying types of data practices that reduce risk—and applying fewer legal constraints to those practices—provides incentives for entities to structure analytic projects that meet lower risk thresholds. Patients are harmed by the failure to learn from clinical data, as well as by the failure to adequately protect it. An effective health data analytics policy framework should effectively address both harms.