Washington and Lee Law Review Online

Volume 72 | Issue 3 Article 3

3-18-2016

Classification Standards for Health Information: Ethical and Practical Approaches

Craig Konnoth
University of Pennsylvania Law School

Classification Standards for Health Information: Ethical and Practical Approaches

Craig Konnoth*

Abstract

Secondary health information research requires vast quantities of data in order to make clinical and health delivery breakthroughs. Restrictive policies that limit the use of such information threaten to stymie this research. While the Notice of Proposed Rulemaking (NPRM) for the new Common Rule permits patients to provide broad consent for the use of their information for research, that policy offers insufficient flexibility. This Article suggests a flexible consenting system that allows patients to consent to a range of privacy risks. The details of the system will be fleshed out in future work.

Table of Contents

* Craig Konnoth, Sharswood Fellow & Lecturer in Law, Senior Fellow, Leonard Davis Institute of Health Economics, University of Pennsylvania Law School.

I. Introduction

Secondary research using health information is on the rise. Not all informational research presents equal burdens. Yet, there has been little commentary on the distinction between different kinds of informational research. This Article helps remedy this problem. In so doing, it sets out the first step towards a blueprint for Institutional Review Boards (IRBs) and other actors who must decide what kinds of constraints to apply to different kinds of information.

The Article first briefly explains what constitutes secondary research of health information and outlines the problem. Second, it points to analogous contexts in which health information is categorized. Third, relying on Helen Nissenbaum's approach to privacy as contextual integrity, it argues for a "scoring" methodology that IRBs should use in determining information sensitivity.

II. Secondary Research of Health Information

Today, medical breakthroughs are increasingly coming from "informational" or "secondary" research, that is, research that aggregates information about patients, including physical conditions, genetic information, treatments, responses, and outcomes. This research gives researchers a real-world snapshot at a population-wide level in a way that is not possible with traditional clinical trials. Data from clinical contexts are fed back into databases in a "continuous feedback loop" that iteratively helps improve clinical and health-delivery outcomes.[1] The new form of research is prominently foregrounded in new policy initiatives. The Affordable Care Act's (ACA) Comparative Effective Research (CER), the Food and Drug Administration's (FDA) Sentinel post-market drug surveillance programs, and the receipt proposed changes to the Common Rule that apply to all

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

federally funded research consciously highlight secondary research approaches.[2]

This new research is important and widespread but requires vast quantities of information. Numerous private payers, major health systems, data intermediaries, and government entities aggregate vast quantities of data that they use and sell to others to, among other purposes, determine health outcomes, marketing practices, health delivery procedures, and the like.[3] For example, agglomerating data has allowed researchers to identify genetic mutations that presage high risks of breast cancer or Alzheimer's.[4] Drug administration in general may change. Drug absorption, drug distribution, drug metabolism and elimination, drug concentration at the target site, and the receptivity of the target receptors may vary from individual to individual based on various factors that secondary research may well discover.[5]

The research also has non-clinical uses. It helps develop necessary health quality measures and helps identify areas to target for cost reduction. Hospital readmission rates, for example, were found to be correlated with mental depression in Washington, D.C. hospitals.[6] This research also supports older mechanisms, such as clinical trials, by helping identify potential subjects that can be targeted for recruitment.[7] And it renders

---

2. *See* Federal Policy for the Protection of Human Subjects, 80 Fed. Reg. 53,933, 53,938 (Sept. 8, 2015) (to be codified at 45 C.F.R. pt. 46) [hereinafter NPRM]; Sharona Hoffman & Andy Podgurski, *Improving Health Care Outcomes Through Personal Comparison of Treatment Effectiveness Based on Electronic Health Records*, 39 J. L. MED. & ETHICS 425, 425 (2011) (noting that CER represents a major public health enterprise, for which Congress has allocated billions of dollars since 2009).

3. *See* Nicholas Terry, *Big Data Proxies and Health Privacy Exceptionalism*, 24 HEALTH MATRIX 65, 66 (2014) (noting that big data facilitates the creation of health data proxies).

4. For a longer list that is screenable through, for example, PGD, see *What We Test For*, GENESIS GENETICS, http://genesisgenetics.org/pgd/what-we-test-for/ (last visited Feb. 15, 2016) (on file with the Washington and Lee Law Review).

5. *See, e.g.*, Allen D. Roses, *Pharmacogenetics and Future Drug Development and Delivery*, 355 LANCET 1358 (2000).

6. ALEX PENTLAND ET AL., BIG DATA AND HEALTH 31 (2013).

7. *See* Tracy Stuardi et al., *Database Recruitment: A Solution to Poor Recruitment in Randomized Trials?*, 28 FAM. PRAC. 329, 329 (2011) (discussing database recruitment); Walter F. Stewart et al., *Bridging the Inferential Gap:*

possible investigations of situations where it would be impossible or unethical to carry out clinical trials, including where doctors suspect a drug has dangerous side effects, or investigating off-label uses. These benefits are just the tip of the iceberg.

The ultimate goal is to create what policymakers call a learning health system.[8] Each clinical intervention will feed back into centrally accessible databases. Analytics from these databases will set standards for treating patients. These standards will optimize treatment protocols after taking into account numerous factors, including the individual genetic and behavioral profile of the patient, and structural considerations, such as cost, limits to medical infrastructure in a particular geographical area, and staffing needs, among other variables. Providers will treat patients depending on their profile using these insights. The outcome from those treatments will be fed back into the database, producing even more refined insights. And of course, data will be used to craft policy decisions and initiatives in areas ranging from drug approval to reimbursement, medical curricular reform, and health discrimination policy.[9]

III. The Problem

Although the fruits of collection are plentiful, information collection imposes burdens on individuals.[10] The inadvertent release of health information can impose objective harms on individuals if it is used inappropriately, ranging from discrimination in employment and insurance, to reputational

---

The Electronic Health Record and Clinical Evidence, 26 HEALTH AFF. 181, 182–83 (2007) (noting the shortcomings of RCTs—that they are too selective and ignore comorbidities—and that secondary research helps bridge the gap).

8. *See* INSTITUTES OF MEDICINE, BEST CARE AT LOWER COST: THE PATH TO CONTINUOUSLY LEARNING HEALTH CARE IN AMERICA 3 (2012) (noting that health care policy improvement requires that payments for services should reward desired care outcomes and movement toward providing the best care at a lower cost).

9. *See generally id.* (explaining how health system analytics can improve the quality of healthcare services).

10. *See generally* sources cited *infra* notes 11–12 (setting forth examples of such burdens).

loss. Even if these harms do not eventually occur, the psychological discomfort that comes from the fear of inappropriate use, whether or not it accurately estimates the risk of these other harms, is itself a separate harm.[11] Finally, even if there is no actual information misuse, or fear of misuse, individuals also consider privacy intrusions as disrespectful and harmful to their autonomy.[12]

Nonetheless, the privacy risks from different kinds of data are different. Existing and proposed research rules seek to take this into account. For example, under the existing Common Rule and under the Health Insurance Portability and Accountability Act (HIPAA), research with sufficiently deidentified information can proceed unhindered. Unless the IRB grants a waiver, researchers need to obtain specific and informed consent before accessing information for secondary research.[13]

The proposed version of the Common Rule does not maintain as strict a division between identified and deidentified data—although HIPAA restrictions would still apply to most secondary data. But it proposes other distinctions. For example, it distinguishes between data collected for research and non-research purposes.[14]

---

11. *See, e.g.*, Ryan Calo, *The Boundaries of Privacy Harm*, 86 IND. L.J. 1131, 1138–39 (2010) (noting the potential harms arising from collection of individuals' health information disclosures).

12. *See* Robert C. Post, *The Social Foundations of Privacy: Community and Self in the Common Law Tort*, 77 CAL. L. REV. 957, 1008 (1989) (noting that the right to control disclosure of intimate information is constitutive of the norms of privacy and human autonomy). These concepts are expanded upon in Craig Konnoth, An Expressive Theory of Privacy Intrusions *passim* (Feb. 10, 2016) (unpublished manuscript) (on file with author).

13. *See generally* 45 C.F.R. § 46.101(b)(4) (2006); SECRETARY'S ADVISORY COMMITTEE ON HUMAN RESEARCH PROTECTIONS (SACHRP), FAQS, TERMS AND RECOMMENDATIONS ON INFORMED CONSENT AND RESEARCH USE OF BIOSPECIMENS (July 20, 2011), http://www.hhs.gov/ohrp/sachrp/commsec/attachmentdfaq'stermsandrecommendations.pdf.pdf (last visited Mar. 9, 2016) (answering frequently asked questions related to relating to informed consent and research use of biospecimens) (on file with the Washington and Lee Law Review).

14. *See* NPRM, *supra* note 2, at 53,933, 53,973 (noting the NPRM's proposal to distinguish between the consent required for data collected for research and non-research purposes).

Yet these distinctions seem to have been picked somewhat arbitrarily. To be sure, it makes a difference whether the information is identifiable and the purposes for which it was collected. But these are only two criteria among a multitude of other criteria that could be utilized. A more systematic method for defining data sensitivity is necessary.

Looking to other arenas provides some assistance. For example, organizations, including the federal government, frequently adopt data classification standards that determine, for example, whether and what data is highly classified or confidential, internally circulable, or publicly available.[15] Yet, these standards gauge the sensitivity of information by the extent to which agency functioning would be impaired if, inter alia, privacy were breached. Because the standards rarely discuss the principles behind the methodology they adopt in ways that can apply to other contexts, they are not of much assistance.

IV. Contextual Integrity and Privacy

In order to develop this account, this Part turns to Helen Nissenbaum's influential explanation of contextual integrity. Privacy involves control over the flow of information. The rules of access and use that ultimately determine flow are determined by context. Nissenbam argues that our lives can generally be divided into multiple contexts, spheres, or fields.[16] Social norms recognized by most members of society as controlling dictate appropriate behavior in those contexts. Communities are defined by these shared norms, "common understandings and shared interests, which . . . facilitate . . . mutual interaction" among their members.[17] The norms of the context may prescribe greater or less access to the information depending on the context. As long as these norms are complied with, there are no privacy intrusions simply because information has been accessed.

---

15. The most important of these is the National Institute of Standards and Technology, *Standards for Security Categorization of Federal Information and Information Systems*, 199 FIPS (2004).

16. *See* HELEN NISSENBAUM, PRIVACY IN CONTEXT 132 (2010) (setting forth this argument).

17. Post, *supra* note 12, at 991.

Thus, for example, I may provide a doctor with certain information that I would never give friends and vice versa. The rules of each context are shaped by numerous forces. In some "highly ritualized" contexts, roles and actions are guided by a detailed set of norms; in other contexts, the norms are less well-defined.[18] There may be variation in the way different individuals or groups treat information if the norms are open-ended or if there is disagreement. Finally, privacy norms are malleable. One can shift norms such that information regarding a particular issue is no longer private in a given context, and public discussion becomes normalized.[19]

To determine whether privacy has been respected, Nissenbaum considers three main, but non-exhaustive, contextual elements:[20] (1) The context, or backdrop, where the disclosure takes place. Contexts are characterized by different kinds of activities and values—such as medical, intimate, educational, and other contexts. The amount and kind of information we circulate will depend on the context. (2) The actors involved, which include the discloser, the disclosee, and the subject of the information. The same context may have different actors with different roles—thus, we may release different kinds and amounts of information to nurses, orderlies, or doctors, all in the medical setting. (3) Transmission principles, which define the kinds of information flow. The release may be forced, voluntary, mutual, unidirectional, etc.[21] To Nissenbaum's account, this Article will add one additional consideration: (4) The purpose of the collection. A doctor reading a patient's chart in a medical setting would only satisfy privacy norms if the *purpose* of her reading the chart was to treat the patient.

---

18. NISSENBAUM, *supra* note 16, at 129.

19. *See, e.g.*, Matt Ferner, *These Photos of Legal Recreational Marijuana Users Shatter Stereotypes*, HUFF. POST (Mar. 16, 2015, 12:32 PM), http://www.huffingtonpost.com/2015/04/16/photos-recreational-marijuana_n_7075710.html (last visited Feb. 15, 2016) ("Marijuana is being covered by the media in an increasingly sophisticated and nuanced way now that the laws are changing and more people are 'out' as marijuana users . . . .") (on file with the Washington and Lee Law Review).

20. NISSENBAUM, *supra* note 16, at 143.

21. *Id.* at 145.

In the case of breach that results in a privacy intrusion, or surveillance, there is a mismatch between the information type and one or more of the remaining contextual elements. In the interests of clarity, this Article shall refer to any element that does not match with the kind of information involved as the mismatched or intruding element. These are the elements that take the place usually occupied by the appropriate element in the particular interaction.

Consider the case of private medical information. If your colleagues seek such information, there will be a mismatch of both context and actors: Medical information generally is not accessed in employment contexts by one's colleagues. Similarly, even in a medical context, the wrong kind of actor—say an orderly—may invade your privacy by reading your medical chart. Next, there could be a mismatch of transmission principles. A doctor can invade your privacy by obtaining your information using the wrong transmission principle—for example, by threatening to withhold treatment unless you volunteer information you otherwise do not wish to. Finally, a mismatch of purpose can result in an intrusion, for example, where a doctor collected medical information in a medical context but for purposes other than treatment—say, to write a research paper without obtaining consent.

Finally, certain circumstances can lessen or increase the risk of a contextual violation. The key consideration is whether the information can be traced back to the individual who provided it. If the information can be traced back, then it becomes that much easier to apply the information to other contexts of the individual's life.

V. Scoring Intrusions

This Article offers Nissenbaum's account as a basis for a scoring system that IRBs should use when determining the type of risk a certain kind of research project presents. The scoring system could consist of various attributes—five of which are listed above. The IRB would score, not contextual variables themselves, but rather, how much the contextual variables of the

proposed research differ from those of the original research—what this Article shall call the differential score.

Thus, if all or some of the original research team are involved in the new research, or if the purpose of the new research does not deviate sharply from the purpose of the old research—for example, because it is examining the same disease—then the differential score would be low. But even if the actors and the purpose are similar, the context may change. The same researchers may have departed from a university and may now be working on the same disease for a commercial institution. This change of context will increase the differential score.

But the IRB must also consider the risk that the information, once conveyed to the new research project, may leak to other contexts. The risk of leakage leading to violations big and small increase dramatically to the degree the data is identified or identifiable. Without identifiable information, it would be that much harder to carry out even more grievous invasions, for example, by transferring information regarding a disease to the employment context. Because the harm that leakage may cause is somewhat incalculable, an IRB may choose not to quantify all of the secondary, or rather, tertiary, harms that could come from leakage of information to yet another context. Rather, they should assess the risk that the information may be re-identified. In so doing, IRBs can draw from existing methodologies used in privacy impact assessments—which themselves have drawn limited academic attention.[22] Using the scoring mechanism, the IRB could come up with a "total score" that would determine what procedures would need to be in place to ensure that the data subjects are projected.

VI. Remedy

With this kind of nuanced scoring system, data subjects should not be protected using an all-or-nothing approach. Under

22. *See* Roger Clarke, *Privacy Impact Assessment: Its Origins and Development*, 25 COMPUTER L. & SECURITY REV. 123, 123 (2009) ("Privacy impact assessment (PIA) is a systematic process for evaluating the potential effects on privacy of a project, initiative or proposed system or scheme.").

the current system, unless an individual consents, or a waiver is granted, identifiable data is out. Under the Common Rule moving forward, if data is not covered by HIPAA rules, there are distinctions between whether the data was collected for research versus non-research purposes, which tracks my suggestion that the *purpose* of the original context and the new context be compared. But overall, identifiable data could be used in certain circumstances if the individual, when the information was collected, offered broad consent for the information to be used for all further research.

Under the new approach, however, individuals can offer a more calibrated breadth of consent. Rather than being offered an all-or-nothing approach—provide broad consent for all future research or only limited consent for this project—they may allow only for research that departs from the original context of collection by a certain amount. The breadth of consent score can be "tagged" to the information.[23] Once an IRB's scoring of a new research project is established, only those records at or above the IRB's score can be (automatically) included in the research.

Problems will remain with research biases. Today, we know that some groups of individuals are categorically less likely to participate in research.[24] Those individuals may also, under the proposed regime, offer narrower consent systematically. Potentially, such biases in studies can be remedied in the following manner. IRBs can, in the right circumstances, apply to Office for Human Research Protections or the relevant federal agency for some sort of minority waiver, under which those individuals in underrepresented groups who offer the broadest consent will also be included in the study in sufficient numbers such that the underrepresentation does not reach a certain threshold. To be sure, that means that some individuals' information will be included in projects to which they would not have consented. But the approach offered by this Article is the

---

23. *Cf.* NPRM *supra* note 2, at 53,973 (suggesting that the information be tagged in various ways to indicated breadth of consent).

24. *See* INSTITUTES OF MEDICINE, CLINICAL DATA AS THE BASIC STAPLE FOR A LEARNING HEALTH SYSTEM 96 (2010) (indicating that people with potentially stigmatizing health conditions, such as those tied to mental health, genetics, or sexually transmitted diseases, are particularly concerned about professional health researchers seeing their medical records).

best approach for achieving the public interest, which sometimes requires overriding private preferences and respecting individual preferences, because those most averse to information collection will not have their data used.

VII. Conclusion

This Article has offered a very brief explanation of a way to reform the manner in which we assess data sensitivity and the way in which we implement protections based on the scoring system. More importantly, it has provided the ethical explanation that undergirds this analysis. The principles offered here can therefore be expanded upon to create an ethical but nuanced and automated system by which to carry out secondary health information research.